Springer Protocols

Methods in Molecular Biology 471

Cancer Epidemiology

Volume 1, Host Susceptibility Factors

Edited by
Mukesh Verma

Humana Press

Cancer Epidemiology

METHODS IN MOLECULAR BIOLOGY™

John M. Walker, SERIES EDITOR

475. Cell Fusion: *Overviews and Methods, edited by Elizabeth H. Chen, 2008*
474. Nanostructure Design: *Methods and Protocols, edited by Ehud Gazit and Ruth Nussinov, 2008*
473. Clinical Epidemiology: *Practice and Methods, edited by Patrick Parfrey and Brendon Barrett, 2008*
472. Cancer Epidemiology, Volume 2: *Modifiable Factors, edited by Mukesh Verma, 2008*
471. Cancer Epidemiology, Volume 1: *Host Susceptibility Factors, edited by Mukesh Verma, 2008*
470. Host-Pathogen Interactions: *Methods and Protocols, edited by Steffen Rupp and Kai Sohn, 2008*
469. Wnt Signaling, Volume 2: *Pathway Models, edited by Elizabeth Vincan, 2008*
468. Wnt Signaling, Volume 1: *Pathway Methods and Mammalian Models, edited by Elizabeth Vincan, 2008*
467. Angiogenesis Protocols: *Second Edition, edited by Stewart Martin and Cliff Murray, 2008*
466. Kidney Research: *Experimental Protocols, edited by Tim D. Hewitson and Gavin J. Becker, 2008*
465. Mycobacteria, Second Edition, edited by *Tanya Parish and Amanda Claire Brown, 2008*
464. The Nucleus, Volume 2: *Physical Properties and Imaging Methods, edited by Ronald Hancock, 2008*
463. The Nucleus, Volume 1: *Nuclei and Subnuclear Components, edited by Ronald Hancock, 2008*
462. Lipid Signaling Protocols, edited by *Banafshe Larijani, Rudiger Woscholski, and Colin A. Rosser, 2008*
461. Molecular Embryology: *Methods and Protocols, Second Edition, edited by Paul Sharpe and Ivor Mason, 2008*
460. Essential Concepts in Toxicogenomics, edited by *Donna L. Mendrick and William B. Mattes, 2008*
459. Prion Protein Protocols, edited by *Andrew F. Hill, 2008*
458. Artificial Neural Networks: Methods and Applications, edited by *David S. Livingstone, 2008*
457. Membrane Trafficking, edited by *Ales Vancura, 2008*
456. Adipose Tissue Protocols, Second Edition, edited by *Kaiping Yang, 2008*
455. Osteoporosis, edited by *Jennifer J. Westendorf, 2008*
454. SARS- and Other Coronaviruses: *Laboratory Protocols*, edited by *Dave Cavanagh, 2008*
453. Bioinformatics, Volume II: *Structure, Function and Applications*, edited by *Jonathan M. Keith, 2008*
452. Bioinformatics, Volume I: *Data, Sequence Analysis and Evolution*, edited by *Jonathan M. Keith, 2008*
451. Plant Virology Protocols: *From Viral Sequence to Protein Function*, edited by *Gary Foster, Elisabeth Johansen, Yiguo Hong, and Peter Nagy, 2008*
450. Germline Stem Cells, edited by *Steven X. Hou and Shree Ram Singh, 2008*
449. Mesenchymal Stem Cells: *Methods and Protocols*, edited by *Darwin J. Prockop, Douglas G. Phinney, and Bruce A. Brunnell, 2008*
448. Pharmacogenomics in Drug Discovery and Development, edited by *Qing Yan, 2008*
447. Alcohol: *Methods and Protocols*, edited by *Laura E. Nagy, 2008*
446. Post-translational Modification of Proteins: *Tools for Functional Proteomics*, Second Edition, edited by *Christoph Kannicht, 2008*
445. Autophagosome and Phagosome, edited by *Vojo Deretic, 2008*
444. Prenatal Diagnosis, edited by *Sinhue Hahn and Laird G. Jackson, 2008*
443. Molecular Modeling of Proteins, edited by *Andreas Kukol, 2008.*
442. RNAi: Design and Application, edited by *Sailen Barik, 2008*
441. Tissue Proteomics: *Pathways, Biomarkers, and Drug Discovery*, edited by *Brian Liu, 2008*
440. Exocytosis and Endocytosis, edited by *Andrei I. Ivanov, 2008*
439. Genomics Protocols, Second Edition, edited by *Mike Starkey and Ramnanth Elaswarapu, 2008*
438. Neural Stem Cells: *Methods and Protocols*, Second Edition, edited by *Leslie P. Weiner, 2008*
437. Drug Delivery Systems, edited by *Kewal K. Jain, 2008*
436. Avian Influenza Virus, edited by *Erica Spackman, 2008*
435. Chromosomal Mutagenesis, edited by *Greg Davis and Kevin J. Kayser, 2008*
434. Gene Therapy Protocols: Volume II: *Design and Characterization of Gene Transfer Vectors*, edited by *Joseph M. LeDoux, 2008*
433. Gene Therapy Protocols: Volume I: *Production and In Vivo Applications of Gene Transfer Vectors*, edited by *Joseph M. LeDoux, 2008*
432. Organelle Proteomics, edited by *Delphine Pflieger and Jean Rossier, 2008*
431. Bacterial Pathogenesis: *Methods and Protocols*, edited by *Frank DeLeo and Michael Otto, 2008*
430. Hematopoietic *Stem Cell Protocols*, edited by *Kevin D. Bunting, 2008*
429. Molecular Beacons: *Signalling Nucleic Acid Probes, Methods and Protocols*, edited by *Andreas Marx and Oliver Seitz, 2008*
428. Clinical Proteomics: *Methods and Protocols*, edited by *Antonia Vlahou, 2008*
427. Plant Embryogenesis, edited by *Maria Fernanda Suarez and Peter Bozhkov, 2008*
426. Structural Proteomics: *High-Throughput Methods*, edited by *Bostjan Kobe, Mitchell Guss, and Huber Thomas, 2008*
425. 2D PAGE: *Sample Preparation and Fractionation*, Volume II, edited by *Anton Posch, 2008*
424. 2D PAGE: *Sample Preparation and Fractionation*, Volume I, edited by *Anton Posch, 2008*
423. Electroporation Protocols: *Preclinical and Clinical Gene Medicine*, edited by *Shulin Li, 2008*
422. Phylogenomics, edited by *William J. Murphy, 2008*
421. Affinity Chromatography: *Methods and Protocols*, Second Edition, edited by *Michael Zachariou, 2008*

METHODS IN MOLECULAR BIOLOGY™

Cancer Epidemiology

Volume I
Host Susceptibility Factors

Edited by

Mukesh Verma, PhD

Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, Maryland

Humana Press

Editor
Mukesh Verma
Division of Cancer Control and Population Sciences
Bethesda, Maryland 20892
USA

Series Editor
John M. Walker
School of Life Sciences
University of Hertfordshire
Hatfield, Hertfordshire, AL10 9AB, UK

ISBN: 978-1-58829-978-1 e-ISBN: 978-1-59745-416-2
ISSN: 1064-3745 e-ISSN: 1940-6029
DOI: 10.1007/978-1-59745-416-2

Library of Congress Control Number: 2008931471

© Humana Press 2009, a part of Springer Science+Business Media, LLC
All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Humana Press, c/o Springer Science + Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.
The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.
While the advice and information in this book are believed to be true and accurate at the date of going to press, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Preface

Population studies facilitate the discovery of genetic and environmental determinants of cancer and the development of new approaches to cancer control and prevention. Furthermore, epidemiology studies play a central role in making health policies. Cancer epidemiology may address a number of research areas such as:

- familial predispositions to colon cancer and breast cancer study to determine whether families who carry a genetic predisposition to breast cancer may also be at risk of colon cancer, and vice versa;
- prospective examination of whether baseline dietary intakes and serum levels of carotenoids and vitamin A are associated with subsequent risk of lung cancer;
- analysis of the relationship between serum levels of sex-steroid hormones and genetic polymorphisms in biosynthesis enzymes in a prospective cohort of pre-menopausal women;
- analysis of the role of HLA-Class II similarity/dissimilarity between sexual partners and the role in HIV transmission, using the multicenter hemophilia cohort study population for the data set;
- multiple comparisons and the effect of stratifying data on study power.

This two-volume set compiles areas of research that cover etiological factors or determinants that contribute in the development of cancer as well as describe the latest technologies in cancer epidemiology. Emphasis is placed on translating clinical observations into interdisciplinary approaches involving clinical, genetic, epidemiologic, statistical, and laboratory methods to define the role of susceptibility genes in cancer etiology; translating molecular genetics advances into evidence-based management strategies (including screening and chemoprevention) for persons at increased genetic risk of cancer; identifying and characterizing phenotypic manifestations of genetic and familial cancer syndromes; counseling individuals at high risk of cancer; investigating genetic polymorphisms as determinants of treatment-related second cancers; and pursuing astute clinical observations of unusual cancer occurrences that might provide new clues to cancer etiology. All the chapters in these two books are divided into three categories:

Volume 1:

Cancer Incidence, Prevalence, Mortality, and Surveillance

Methods, Technologies, and Study Design in Cancer Epidemiology

Host Susceptibility Factors in Cancer Epidemiology

Volume 2:

Modifiable Factors in Cancer Epidemiology

Epidemiology of Organ-Specific Cancer

These chapters have been written in a way that allows readers to get the maximum advantage of the methods involved in cancer epidemiology. Several examples of specific organ sites would be helpful in understanding cancer etiology.

Mukesh Verma, Ph.D.

Contents

Preface . . . . . *v*
Contributors . . . . . *ix*
Contents of Volume II . . . . . *xi*

PART I: CANCER INCIDENCE, PREVALENCE, MORTALITY AND SURVEILLANCE

1. Cancer Occurrence 3
Ahmedin Jemal, Melissa M. Center, Elizabeth Ward, and Michael J. Thun
2. Cancer Registry Databases: *An Overview of Techniques of Statistical Analysis and Impact on Cancer Epidemiology* 31
Ananya Das
3. Breast Cancer in Asia 51
Cheng-Har Yip
4. Cancer Epidemiology in the United States: Racial, Social, and Economic Factors 65
Dana Sloane
5. Epidemiology of Multiple Primary Cancers 85
Isabelle Soerjomataram and Jan Willem Coebergh
6. Cancer Screenings, Diagnostic Technology Evolution, and Cancer Control 107
Fabrizio Stracci
7. Thriving for Clues in Variations seen in Mortality and Incidence of Cancer: *Geographic Patterns, Temporal Trends, and Human Population Diversities in Cancer Incidence and Mortality* 137
Alireza Mosavi-Jarrahi and Mohammad Ali Mohagheghi

PART II: METHODS, TECHNOLOGIES AND STUDY DESIGN IN CANCER EPIDEMIOLOGY

8. Evaluation of Environmental and Personal Susceptibility Characteristics That Modify Genetic Risks 163
Jing Shen
9. Introduction to the Use of Regression Models in Epidemiology 179
Ralf Bender
10. Proteomics and Cancer Epidemiology 197
Mukesh Verma
11. Different Study Designs in the Epidemiology of Cancer: *Case-Control vs. Cohort Studies* 217
Harminder Singh and Salaheddin M. Mahmud
12. Methods and Approaches in Using Secondary Data Sources to Study Race and Ethnicity Factors 227
Sujha Subramanian
13. Statistical Methods in Cancer Epidemiologic Studies 239
Xiaonan Xue and Donald R. Hoover

14. Methods in Cancer Epigenetics and Epidemiology 273
Deepak Kumar and Mukesh Verma

PART III: HOST SUSCEPTIBILITY FACTORS IN CANCER EPIDEMIOLOGY

15. Mitochondrial DNA Polymorphism and Risk of Cancer 291
Keshav K. Singh and Mariola Kulawiec

16. Polymorphisms of DNA Repair Genes: *ADPRT, XRCC1 and XPD and Cancer Risk in Genetic Epidemiology* 305
Jun Jiang, Xiuqing Zhang, Huanming Yang and Wendy Wang

17. Risk Factors and Gene Expression in Esophageal Cancer 335
Xiao-chun Xu

18. Single Nucleotide Polymorphisms in DNA Repair Genes and Prostate Cancer Risk 361
Jong Y. Park, Yifan Huang and Thomas A. Sellers

19. Linking the Kaposi's Sarcoma-Associated Herpesvirus (KSHV/HHV-8) to Human Malignancies 387
Inna Kalt, Shiri-Rivka Masa and Ronit Sarid

20. Cancer Cohort Consortium Approach: *Cancer Epidemiology in Immunosuppressed Groups* 409
Diego Serraino, Pierluca Piselli for the Study Group

21. Do Viruses Cause Breast Cancer? 421
James S. Lawson

22. Epidemiology of Human Papilloma Virus (HPV) in Cervical Mucosa 439
Subhash C. Chauhan, Meena Jaggi, Maria C. Bell, Mukesh Verma and Deepak Kumar

23. Epigenetic Targets in Cancer Epidemiology 457
Ramona G. Dumitrescu

24. Epidemiology of Lung Cancer Prognosis: *Quantity and Quality of Life* 469
Ping Yang

25. Hereditary Breast and Ovarian Cancer Syndrome: *The Impact of Race on Uptake of Genetic Counseling and Testing* 487
Michael S. Simon and Nancie Petrucelli

Index *501*

Contributors

Maria C. Bell • *Cancer Biology Research Center, Sanford Research/USD and Department of Obstetrics and Gynecology, Sanford School of Medicine, The University of South Dakota, Sioux Falls, SD, USA*

Ralf Bender • *Institute for Quality and Efficiency in Health Care, Cologne, Germany*

Melissa M. Center • *Department of Epidemiology and Surveillance Research, American Cancer Society, Atlanta, GA, USA*

Subhash C. Chauhan • *Cancer Biology Research Center, Sanford Research/USD and Department of Obstetrics and Gynecology, Sanford School of Medicine, The University of South Dakota, Sioux Falls, SD, USA*

Jan Willem Coebergh • *Comprehensive Cancer Centre South, AE Eindhoven, The Netherlands*

Ananya Das • *Department of Medicine, Mayo College of Medicine, Mayo Clinic, Scottsdale, AZ, USA*

Ramona G. Dumitrescu • *Georgetown University Medical Center, Lombardi Cancer Center, Washington, DC, USA*

Donald R. Hoover • *Department of Statistics and Institute for Health Care Policy and Aging Research, Rutgers University, New Brunswick, NJ, USA*

Yifan Huang • *Division of Cancer Prevention and Control, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA*

Meena Jaggi • *Cancer Biology Research Center, Sanford Research/USD and Dept of Obstetrics and Gynecology, Sanford School of Medicine, The University of South Dakota, Sioux Falls, SD, USA*

Ahmedin Jemal • *Department of Epidemiology and Surveillance Research, American Cancer Society, Atlanta, GA, USA*

Jun Jiang • *Beijing Institute of Genomics, Chinese Academy of Sciences (BIG, CAS), China*

Inna Kalt • *The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat Gan, Israel*

Mariola Kulawiec • *Department of Cancer Genetics, Roswell Park Cancer Institute, Buffalo, NY, USA*

Deepak Kumar • *Department of Biological and Environmental Sciences, University of the District of Columbia, Washington, DC, USA*

James S. Lawson • *School of Public Health, University of New South Wales, Sydney, Australia*

Salaheddin M. Mahmud • *Department of Community Health Sciences. University of Manitoba, Winnipeg, MB, Canada*

Shiri-Rivka Masa • *The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat Gan, Israel*

Mohammad Ali Mohagheghi • *The Cancer Research Center of the Cancer Institute, Tehran, Iran*

Alireza Mosavi-Jarrahi • *Dept. Of Epidemiology, School of Public Health, Shaheed Beheshti University of Medical Sciences, and The Cancer Research Center of the Cancer Institute, Tehran, Iran*

Jong Y. Park • *Division of Cancer Prevention and Control, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA*

Nancie Petrucelli • *Population Studies and Prevention Program, Karmanos Cancer Institute at Wayne State University, Detroit, MI, USA*
Pierluca Piselli • *Epidemiology Unit, National Institute for Infectious Diseases, IRCCS L. Spallanzani, Rome, Italy*
Ronit Sarid • *The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat Gan, Israel*
Thomas A. Sellers • *Division of Cancer Prevention and Control, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA*
Diego Serraino • *Virology Section, University Campus Bio-Medico, Spallanzani Hospital, Rome, Italy*
Jing Shen • *Dept of Environmental Health Sciences, Columbia University, New York, NY, USA*
Michael S. Simon • *Division of Hematology and Oncology, Karmanos Cancer Institute at Wayne State University, Detroit, MI, USA*
Harminder Singh • *Department of Medicine, University of Manitoba, Winnipeg, MB, Canada*
Keshav K. Singh • *Department of Cancer Genetics, Roswell Park Cancer Institute, Buffalo, NY, USA*
Dana A. Sloane • *Division of Gastroenterology, Washington Hospital Center, Washington, DC, USA*
Isabelle Soerjomataram • *Department of Public Health, Erasmus MC, Rotterdam, The Netherlands*
Fabrizio Stracci • *Department of Surgical and Medical Specialties, and Public Health, University of Perugia, Perugia, Italy*
Sujha Subramanian • *RTI International, Waltham, MA, USA*
Michael J. Thun • *Department of Epidemiology and Surveillance Research, American Cancer Society, Atlanta, GA, USA*
Mukesh Verma • *National Cancer Institute, Bethesda, MD, USA*
Wendy Wang • *Division of Cancer Prevention, National Cancer Institute, National Institute of Health (DCP/NCI/NIH), USA*
Elizabeth Ward • *Department of Epidemiology and Surveillance Research, American Cancer Society, Atlanta, GA, USA*
Xiuqing Zhang • *Beijing Institute of Genomics, Chinese Academy of Sciences (BIG, CAS), Beijing, China*
Xiao-chun Xu • *Department of Clinical Cancer Prevention, The University of Texas M. D. Anderson Cancer Center, Houston, TX, USA*
Xiaonan Xue • *Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY, USA*
Huanming Yang • *Beijing Institute of Genomics, Chinese Academy of Sciences (BIG, CAS), China*
Ping Yang • *Mayo Clinic, College of Medicine, Rochester, MN, USA*
Cheng-Har Yip • *Department of Surgery, University Malaya, Medical Centre, Kuala Lumpur, Malaysia*

Contents of Volume II

PART I: MODIFIABLE FACTORS IN CANCER EPIDEMIOLOGY

1. Environmental and Occupational Risk Factors for Lung Cancer
   Irene Brüske-Hohlfeld
2. Lifestyle, Genes, and Cancer
   Yvonne M. Coyle
3. Energy Balance, Physical Activity, and Cancer Risk
   Alecia Malin Fair and Kara Montgomery
4. Genetic Epidemiology Studies in Hereditary Non-Polyposis Colorectal Cancer
   Rodney J. Scott and Jan Lubinski
5. Parental Smoking and Childhood Leukemia
   Jeffrey S. Chang
6. Lung Cancer and Exposure to Metals: *The Epidemiological Evidence*
   Pascal Wild, Eve Bourgkard, and Christophe Paris
7. Breast Cancer and the Role of Exercise in Women
   Beverly S. Reigle and Karen Wonders
8. Energy Intake, Physical Activity, Energy Balance, and Cancer: *Epidemiologic Evidence*
   Sai Yi Pan and Marie DesMeules
9. Contribution of Alcohol and Tobacco Use in Gastrointestinal Cancer Development
   Helmut K. Seitz and Chin Hin Cho
10. Role of Xenobiotic Metabolic Enzymes in Cancer Epidemiology
    Madhu S. Singh and Michael Michael
11. Genetic Polymorphisms in the Transforming Growth Factor-β Signaling Pathways and Breast Cancer Risk and Survival
    Wei Zheng

PART II: EPIDEMIOLOGY OF ORGAN-SPECIFIC CANCER

12. Molecular Epidemiology of DNA Repair Genes in Bladder Cancer
    Anne E. Kiltie
13. Breast Cancer Screening and Biomarkers
    Mai Brooks
14. Epidemiology of Brain Tumors
    Hiroko Ohgaki
15. Mammographic Density: A Heritable Risk Factor for Breast Cancer
    N. F Boyd, L. J. Martin, J. M. Rommens, and A. D. Paterson, S. Minkin, M. J. Yaffe, J. Stone, and J. L. Hopper
16. Acquired Risk Factors for Colorectal Cancer
    Otto S. Lin

17. Aberrant Crypt Foci in Colon Cancer Epidemiology
Sharad Khare, Kamran Chaudhary, Marc Bissonnette, and Robert Carroll

18. Determinants of Incidence of Primary Fallopian Tube Carcinoma (PFTC)
Annika Riska and Arto Leminen

19. The Changing Epidemiology of Lung Cancer
Chee-Keong Toh

20. Epidemiology of Ovarian Cancer
Jennifer Permuth-Wey and Thomas A. Sellers

21. Epidemiology, Pathology, and Genetics of Prostate Cancer Among African Americans Compared with Other Ethnicities
Heinric Williams and Isaac J. Powell

22. Racial Differences in Clinical Outcome After Prostate Cancer Treatment
Takashi Fukagai, Thomas Namiki, Robert G. Carlile, and Mikio Namiki

23. Epidemiology of Stomach Cancer
Hermann Brenner, Dietrich Rothenbacher, and Volker Arndt

Index

Chapter 1

Cancer Occurrence

Ahmedin Jemal, Melissa M. Center, Elizabeth Ward, and Michael J. Thun

Summary

In this chapter, we describe the variability of cancer occurrence by using measures of incidence, mortality, prevalence, and survival, according to demographic characteristics such as age, sex, socioeconomic status, and race/ethnicity, as well as geographic location and time period. We also discuss the variability of cancer occurrence in relation to changes in risk factors, screening rates, and improved treatments. The variation according to risk factors provides strong evidence that much of cancer is caused by environmental factors and is potentially avoidable.

Key words: Cancer incidence, life style factors, mortlaity rate, prevalence, surveillance, survivorship.

1. Introduction

The occurrence of cancer varies substantially according to age, sex, socioeconomic status (SES), race/ethnicity, geographic location, and time period. These variations have provide strong evidence that much of cancer is caused by environmental factors and is potentially avoidable *(1)*. Monitoring the occurrence of cancer in relation to demographic characteristics, geographic location, and time is also important to assess the effectiveness of cancer control efforts in the overall population and in subgroups that may be at higher risk.

Mukesh Verma (ed.), *Methods in Molecular Biology, Cancer Epidemiology, vol. 471*
© 2009 Humana Press, a part of Springer Science + Business Media, Totowa, NJ
Book doi: 10.1007/978-1-59745-416-2

2. Data Sources and Measurements

2.1. Incidence and Mortality

Incidence and mortality are two frequently used measures of cancer occurrence. These indices quantify the number of new cancer cases or deaths, respectively, in a defined population over a specified time period, usually 1 year. Incidence and death rates are commonly expressed as counts per 100,000 people per year. Incidence data are collected by regional and national population-based cancer registries. Population-based cancer registries collect data from all medical facilities, including hospitals, doctors' offices, radiation facilities, and diagnostic laboratories. The type of cancer is coded according to the International Classification of Diseases for Oncology, which assigns an anatomic site and a histologic code *(2)*. In the United States, the National Cancer Institute has monitored the Surveillance, Epidemiology, and End Results (SEER) Program since 1973. Within SEER, nine population-based cancer registries, covering approximately 10% of the U.S. population, have collected information on demographic characteristics of new cancer patients, anatomic site of the tumor, histologic type, extent of disease at the time of diagnosis, first course of treatment, and follow-up for vital status. These nine registries provide the best information on temporal trends in cancer incidence over the past three decades in the United States. Over time, the SEER Program has been expanded so that it now covers approximately 26% of the U.S. population (http://www.seer.cancer.gov).

The Centers for Disease Control and Prevention established the National Program of Cancer Registries (NPCR) in 1994 to expand the geographic coverage of cancer registries. Goals of the program are to improve existing non-SEER population-based cancer registries and to establish new statewide cancer registries (http://www.cdc.gov/cancer/npcr). NPCR currently funds central cancer registries in 45 states, the District of Columbia, and the territories of Puerto Rico, the Republic of Palau, and the Virgin Islands. Through the NPCR and SEER programs, data on cancer incidence are collected for >95% of the U.S. population *(3)*.

Mortality data have been collected for most of the United States since 1930, based on information from death certificates. The underlying cause of death is classified according to the most current International Statistical Classification of Diseases (ICD) *(4)*. Deaths that occurred since 1999 in the United States are classified according to ICD-10, replacing ICD-9, used from 1979 to 1998. The classification of deaths due to cancer is considered accurate in the United States, with approximately 90% agreement between death certificates and pathology reports *(5)*. Mortality

data are available from the National Center for Health Statistics (http://www.cdc.gov/nchs/nvss.html).

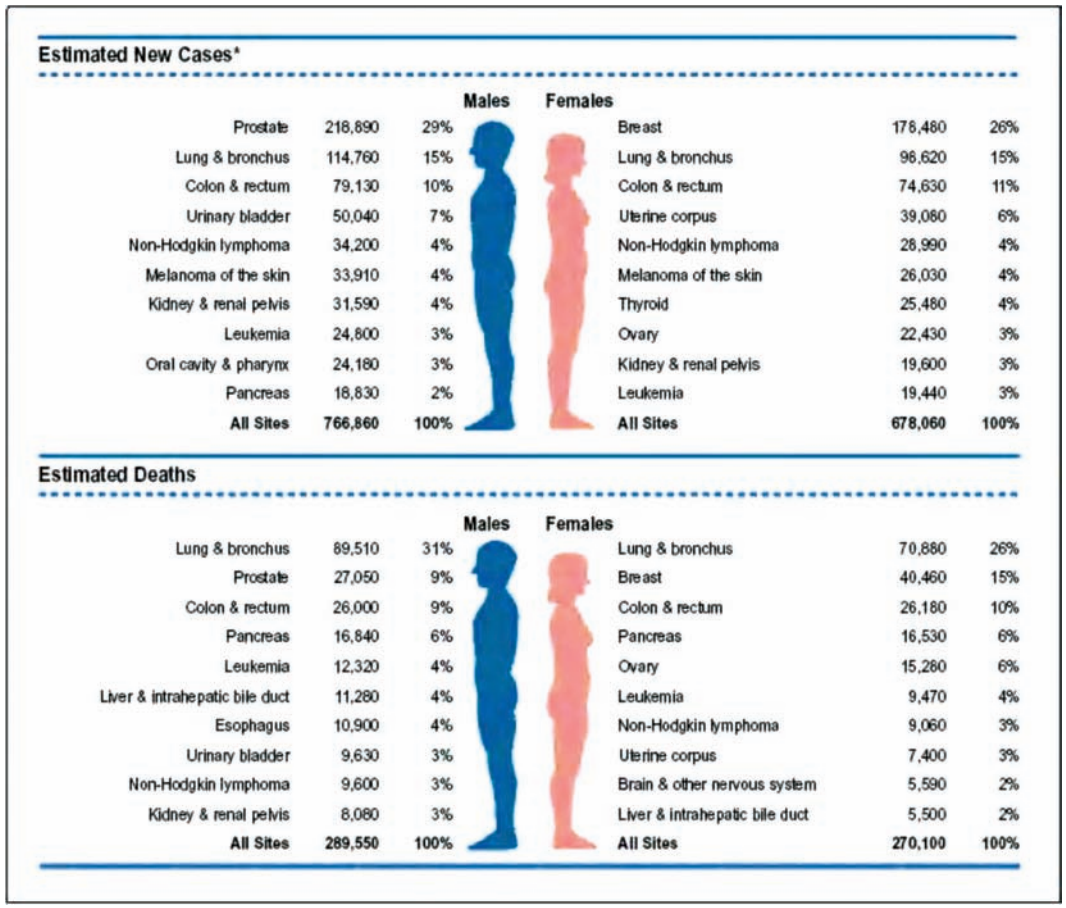
The International Agency for Research on Cancer (IARC; http://www.iarc.fr) compiles worldwide data on cancer incidence, mortality, and prevalence. This information can be accessed through the GLOBOCAN database *(6)*. Only regional incidence data are collected in most countries because very few countries have National cancer registration. For countries with no cancer registry (the case in most developing countries), incidence data are obtained by converting mortality data or by borrowing information from neighboring countries *(7)*. A limitation of this method is that it assumes equal survival across neighboring countries.

2.2. Number of New Cancer Cases and Deaths

Another measure of the cancer burden in a population is the total number of new cases and deaths that occur in a given year in a specified community, irrespective of its size or age distribution. This measure reflects the absolute number of new cases that must be cared for by medical providers and social services. However, actual data on the number of new cancer cases and deaths in the current year are not available due to time required for collection and compilation of information, which takes about 3–4 years. Therefore, the American Cancer Society provides estimates of the total number of new cancer cases and deaths that will occur nationally in the United States and in each state in the current year (**Fig. 1.1**). This is achieved by modeling historic information on the observed cancer incidence and mortality in past years and projecting trends in the number of cancer cases and deaths over time *(8,9)*. About 1.4 million new cancer cases and 556,000 deaths are expected to occur in 2007 in the United States. The number of new cancer cases diagnosed in each year increases with growth and aging of the population. This number will double in approximately 50 years if current age-adjusted incidence rates remain unchanged, simply because of changes in the population *(10)*. Despite the aging and growth of the population, however, the total number of actual recorded cancer deaths in the United States decreased for two consecutive years (2003 and 2004) after increasing for >70 years. The reduction in cancer deaths and death rates is attributed to prevention, early detection, and improved treatment. Whether the total number of actual recorded cancer deaths will continue to decrease despite aging of the post-World War II generation is unknown.

2.3. Prevalence

Whereas incidence rates measure the rate at which new cancer cases are diagnosed in a specified population and time interval, prevalence measures the proportion of people living with cancer at a certain point in time. In principle, the number of prevalent cases includes newly diagnosed cases, those who are under treatment,



Estimated New Cases*

Males			Females		
Prostate	218,890	29%	Breast	178,480	26%
Lung & bronchus	114,760	15%	Lung & bronchus	96,620	15%
Colon & rectum	79,130	10%	Colon & rectum	74,630	11%
Urinary bladder	50,040	7%	Uterine corpus	39,080	6%
Non-Hodgkin lymphoma	34,200	4%	Non-Hodgkin lymphoma	28,990	4%
Melanoma of the skin	33,910	4%	Melanoma of the skin	26,030	4%
Kidney & renal pelvis	31,590	4%	Thyroid	25,480	4%
Leukemia	24,600	3%	Ovary	22,430	3%
Oral cavity & pharynx	24,180	3%	Kidney & renal pelvis	19,600	3%
Pancreas	18,830	2%	Leukemia	19,440	3%
All Sites	**766,860**	**100%**	**All Sites**	**678,060**	**100%**

Estimated Deaths

Males			Females		
Lung & bronchus	89,510	31%	Lung & bronchus	70,880	26%
Prostate	27,050	9%	Breast	40,460	15%
Colon & rectum	26,000	9%	Colon & rectum	26,180	10%
Pancreas	16,840	6%	Pancreas	16,530	6%
Leukemia	12,320	4%	Ovary	15,280	6%
Liver & intrahepatic bile duct	11,280	4%	Leukemia	9,470	4%
Esophagus	10,900	4%	Non-Hodgkin lymphoma	9,060	3%
Urinary bladder	9,630	3%	Uterine corpus	7,400	3%
Non-Hodgkin lymphoma	9,600	3%	Brain & other nervous system	5,590	2%
Kidney & renal pelvis	8,080	3%	Liver & intrahepatic bile duct	5,500	2%
All Sites	**289,550**	**100%**	**All Sites**	**270,100**	**100%**

Fig. 1.1. Leading sites of new cancer cases and deaths—2007 estimates. * Excludes basal and squamous cell skin cancer and *in situ* carcinomas except urinary bladder. Percentage may not total 100% (*CA Cancer J Clin* 57, 43–66, 2007).

and people who are in remission. It is influenced by both the incidence rate of the cancers of interest and by survival or cure rates. In practice, estimates of the number of prevalent cases cannot distinguish precisely between people who have been cured and those with active disease. Nevertheless, prevalence is a useful measure of the health care burden for planning future medical care needs.

Based on extrapolated data from the nine oldest SEER cancer registries, approximately 10.7 million people living in the United States on January 1, 2004 had been diagnosed with cancer *(11)*. This number is expected to increase over time because of improvements in survival and the anticipated growth and aging of the population. The National Cancer Institute also provides estimates on partial and complete prevalence of cancer in the United States. Partial prevalence (limited duration prevalence) is the number of people alive at the prevalence date who had a

diagnosis of the disease within a given number of years in the past, whereas complete prevalence refers to those with a history of diagnosis irrespective of when diagnosis occurred. According to the National Cancer Institute estimates, there were more prevalent cancer cases in women (5.9 million) than in men (4.8 million) in 2004. This difference largely reflects high survival rates of women with early-stage breast cancers and the greater longevity of women than men.

As mentioned, the number of prevalent cases of specific cancer varies according to incidence and survival rates for that cancer. Prostate and breast cancers are the most prevalent cancers in men and women, respectively (**Table 1.1**). The number of prevalent cases of thyroid cancer, which is uncommon but highly curable, is greater than the number of prevalent cases of pancreatic cancer, which is much more common but the most fatal of all cancers.

2.4. Probability of Developing Cancer

The probability that an individual will develop or die from cancer by a certain age is another measure used to describe average risk in the general population (**Table 1.2**). The probability, usually expressed as percentage, also can be expressed as one person in

Table 1.1
Estimated number of prevalent cases of selected cancers, by sex, United States, 2003

	Male		Female		Total	
	No.	%	No.	%	No.	%
All sites	4,692,382	100	5,803,603	100	10,495,985	100
Breast	12,241	0.3	2,356,795	40.6	2,369,036	22.6
Cervix			253,781	4.4	253,781	2.4
Colon and rectum	514,789	11.0	553,409	9.5	1,068,198	10.2
Esophagus	18,104	0.4	5,961	0.1	24,065	0.2
Leukemia	112,324	2.4	86,689	1.5	199,013	1.9
Lung and bronchus	173,431	3.7	181,558	3.1	354,989	3.4
Melanoma	320,178	6.8	342,255	5.9	662,433	6.3
Pancreas	13,302	0.3	14,386	0.2	27,688	0.3
Prostate	1,937,798	41.3			1,937,798	18.5
Thyroid	79,275	1.7	268,149	4.6	347,424	3.3

Source: Surveillance, Epidemiology, and End Results Program, 1975–2003, Division of Cancer Control and Population Sciences, National Cancer Institute, 2006.

Table 1.2
Probability (percentage) of developing invasive cancers over specified age intervals, by sex, United States[a]

		Birth–39 (%)	40–59 (%)	60–69 (%)	70 (%)	Birth to death (%)
All sites[b]	Male	1.42 (1 in 70)	8.69 (1 in 12)	16.58 (1 in 6)	39.44 (1 in 3)	45.31 (1 in 2)
	Female	2.03 (1 in 49)	9.09 (1 in 11)	10.57 (1 in 9)	26.60 (1 in 4)	37.86 (1 in 3)
Bladder[c]	Male	0.02 (1 in 4,381)	0.41 (1 in 241)	0.96 (1 in 105)	3.41 (1 in 29)	3.61 (1 in 28)
	Female	0.01 (1 in 9,527)	0.13 (1 in 782)	0.26 (1 in 379)	0.96 (1 in 105)	1.14 (1 in 87)
Breast	Female	0.48 (1 in 210)	3.98 (1 in 25)	3.65 (1 in 27)	6.84 (1 in 15)	12.67 (1 in 8)
	Male	0.07 (1 in 1,342)	0.93 (1 in 107)	1.67 (1 in 60)	4.92 (1 in 20)	5.79 (1 in 17)
Colon and rectum	Female	0.07 (1 in 1,469)	0.73 (1 in 138)	1.16 (1 in 863)	4.45 (1 in 22)	5.37 (1 in 19)
	Male	0.16 (1 in 640)	0.22 (1 in 452)	0.35 (1 in 286)	1.17 (1 in 86)	1.49 (1 in 67)
Leukemia	Female	0.12 (1 in 820)	0.14 (1 in 694)	0.20 (1 in 491)	0.75 (1 in 132)	1.05 (1 in 95)
	Male	0.03 (1 in 3,146)	1.09 (1 in 92)	2.61 (1 in 38)	6.76 (1 in 15)	8.02 (1 in 12)
Lung and bronchus	Female	0.04 (1 in 2,779)	0.85 (1 in 117)	1.84 (1 in 54)	4.52 (1 in 22)	6.15 (1 in 16)
	Male	0.13 (1 in 775)	0.53 (1 in 187)	0.56 (1 in 178)	1.32 (1 in 76)	2.04 (1 in 49)
Melanoma of the skin	Female	0.21 (1 in 467)	0.42 (1 in 237)	0.29 (1 in 347)	0.62 (1 in 163)	1.38 (1 in 73)
	Male	0.14 (1 in 735)	0.45 (1 in 222)	0.57 (1 in 176)	1.56 (1 in 64)	2.14 (1 in 47)
Non-Hodg–kin lymphoma	Female	0.08 (1 in 1,200)	0.32 (1 in 313)	0.44 (1 in 229)	1.30 (1 in 77)	1.83 (1 in 55)
Prostate	Male	0.01 (1 in 10,373)	2.59 (1 in 39)	7.03 (1 in 14)	13.83 (1 in 7)	17.12 (1 in 6)
Uterine cervix	Female	0.16 (1 in 631)	0.29 (1 in 346)	0.14 (1 in 695)	0.20 (1 in 512)	0.73 (1 in 138)
Uterine corpus	Female	0.06 (1 in 1,652)	0.70 (1 in 142)	0.81 (1 in 124)	1.28 (1 in 78)	2.49 (1 in 40)

[a] For persons free of cancer at beginning of age interval. Based on cancer cases diagnosed during 2001 to 2003. The "1 in" statistics and the inverse of the percentage may not be equivalent due to rounding.). [b] All sites exclude basal and squamous cell skin cancer and *in situ* cancers except urinary bladder.). [c] Includes invasive and *in situ* cancer cases.).
Source: DEVCAN Software, Probability of Developing or Dying of Cancer Software, Version 6.1.0. Statistical Research and Application Branch, National Cancer Institute, 2006. (http://srab.cancer.gov/devcan).

X persons. For example, the lifetime risk of developing lung cancer in U.S. men, 8.02%, can be expressed as 1 in 12 men developing lung cancer in his lifetime. These estimates are based on the average experience of the general population and they may over- or underestimate individual risk because of family history or individual risk factors. For example, the estimate that lung cancer will develop in 1 of 12 men over a lifetime overestimates the risk for nonsmokers and underestimates the risk for smokers. Software to estimate probability of developing or dying from cancer within specified age ranges is available at http://srab.cancer.gov/devcan.

2.5. Survival Rates

The survival rate reflects the proportion of people alive at a specified period after diagnosis, usually 5 years. Survival is commonly expressed as observed or relative. The former rate quantifies the proportion of cancer patients alive after 5 years of follow-up since diagnosis, taking into account deaths from any condition. In contrast, relative survival rates compare the proportion of cancer patients alive 5 years after diagnosis to the corresponding proportions in persons of the same age and sex without cancer. Thus, the relative survival rate reflects the specific effect of the cancer on shortened survival. For example, the 89% 5-year relative survival rate for breast cancer translates to 14% fewer breast cancer patients surviving for 5 years compared with their peers in the general population. Whereas the observed survival rate is commonly used in evaluating the effectiveness of new cancer treatments in age-matched randomized clinical trials, relative survival rates are used to measure progress in early detection and cancer treatment in the general population.

Table 1.3 shows 5-year relative survival rates for selected cancer sites in patients diagnosed during two time intervals, from 1975 to 1977 and from 1996 to 2002. For all races combined,

Table 1.3
Changes in 5-year relative survival rates[a] (%), by race and year of diagnosis in the United States from 1975 to 2002

	All races			White			African American		
Site	1975–1977	1996–2002	Dif.	1975–1977	1996–2002	Dif.	1975–1977	1996–2002	Dif.
All cancers	50	66	16[b]	51	68	17[b]	40	57	17[b]
Brain	24	34	10[b]	23	34	11[b]	26	37	11[b]
Breast (female)	75	89	14[b]	76	90	14[b]	63	77	14[b]
Colon	51	65	14[b]	52	66	14[b]	46	54	8[b]

(continued)

Table 1.3 (continued)

Site	All races			White			African American		
	1975–1977	1996–2002	Dif.	1975–1977	1996–2002	Dif.	1975–1977	1996–2002	Dif.
Esophagus	5	16	11[b]	6	17	11[b]	3	12	9[b]
Hodgkin's disease	73	86	13[b]	74	87	13[b]	71	81	10[b]
Kidney	51	66	15[b]	51	66	15[b]	50	66	16[b]
Larynx	66	65	–1	67	67	0	59	52	–7
Leukemia	35	49	14[b]	36	50	14[b]	33	39	6
Liver and bile duct	4	10	6[b]	4	10	6[b]	2	7	5[b]
Lung and bronchus	13	16	3[b]	13	16	3[b]	12	13	1[b]
Melanoma of the skin	82	92	10[b]	82	93	11[b]	58	75	17
Multiple myeloma	26	33	7[b]	25	33	8[b]	31	32	1
Non-Hodgkin lymphoma	48	63	15[b]	48	64	16[b]	48	56	8
Oral cavity	53	60	7[b]	55	62	7[b]	36	40	4
Ovary[c]	37	45	8[b]	36	45	9[b]	43	39	–4
Pancreas	2	5	3[b]	3	5	2[b]	2	5	3[b]
Prostate	69	100	31[b]	70	100	30[b]	61	98	37[b]
Rectum	49	66	17[b]	49	66	17[b]	45	59	14[b]
Stomach	16	24	8[b]	15	22	7[b]	16	23	7[b]
Testis	83	96	13[b]	83	96	13[b]	82	89	7
Thyroid	93	97	4[b]	93	97	4[b]	91	94	3
Urinary bladder	73	82	9[b]	74	83	9[b]	50	65	15[b]
Uterine cervix	70	73	3[b]	71	75	4[b]	65	66	1
Uterine corpus	87	84	–3[b]	89	86	–3[b]	61	61	0

[a]Survival rates are adjusted for normal life expectancy and are based in cases diagnosed from 1975 to 1977 and 1996 to 2002 and followed through 2003.[b]The difference in rates between 1975 and 1977 and 1996 and 2002 is statistically significant ($p < 0.05$).[c]Recent changes in classification of ovarian cancer, namely, excluding borderline tumors, has affected 1996–2002 survival rates. Note: All cancers excludes basal and squamous cell skin cancers and in situ carcinomas except urinary bladder.Source: Surveillance, Epidemiology, and End Results Program, 1973–2000, Division of Cancer Control and Population Sciences, National Cancer Institute, 2003.

the 5-year survival rate for cancer patients diagnosed during 1996 to 2002 varied from 5% for pancreatic cancer to 100% for prostate cancer. Survival rates increased over the two periods for all cancer types except for corpus uterine and larynx, with prostate cancer showing the highest increase (31%).

The observed increases in 5-year survival reflect a complex mixture of actual improvements in prognosis and artifactual effects of screening. Ideally, early detection leads to a more favorable prognosis and improved survival because treatment given before onset of clinical symptoms is more effective than treatment after, when the disease is at a more advanced stage. At the population level, this translates to true reductions in mortality. However, early detection can increase survival in the absence of effective treatments by merely advancing the time of diagnosis. This is often referred as lead-time bias, with "lead time" referring to the interval between when the cancer is diagnosed because of screening and when it would have been diagnosed clinically. Screening also may increase survival by changing the case mix, that is, by detecting tumors that would not otherwise have been detected during the person's lifetime. For example, the increase in 5-year relative survival rate from 69% for prostate cancer patients diagnosed in 1975 to 1977 to 100% for those diagnosed in 1996 to 2002 in part reflects the detection of tumors that might otherwise have remained clinically inapparent, but they were discovered because of prostate-specific antigen (PSA) testing.

The overall 5-year relative survival rate increased approximately equally in whites and blacks **(Table 1.3)**. However, survival rates in the most recent time period are still substantially lower among blacks than whites, reflecting later stage at diagnosis and poorer stage-specific survival among blacks compared with whites. These differences are thought to largely reflect economic, structural, and cultural barriers to high-quality medical care for the black population. Poorer survival also has been observed for members of other racial and ethnic minority groups. Relative survival rates cannot be calculated for other groups because reliable life tables do not exist to predict survival in the general population. However, an analysis of cause-specific survival rate for patients in the SEER registry diagnosed between 1988 and 1997 found that, in general, all minority populations (Hispanic whites, blacks, American Indians/Alaskan natives, and Hawaiian natives) have an increased probability of dying from cancer within 5 years of diagnosis compared with non-Hispanic whites diagnosed at the same age and stage *(12)*.

2.6. Age Standardization

Because age profoundly affects cancer rates (described below), age standardization is widely used to summarize age-specific rates for comparisons of incidence and death rates between two or more populations with differing age composition. This is achieved by

applying the age-specific rates in the populations of interest to a standard set of weights based on a common age distribution. Standardization removes the effect of the differences in age in the populations being compared. The standardized rate is a hypothetical rate that would be observed in each population were the age composition the same as in the standard population.

Beginning with cancer cases and deaths that occurred in 1999, incidence and mortality rates in the United States have been standardized to the 2000 U.S. standard population *(13)*. Previously, rates were standardized to the 1970 standard population. The purpose of shifting from the 1970 standard to the 2000 standard was to approximate the current age distribution of the U.S. population so that the age-standardized rates would more closely resemble contemporary incidence and death rates. The new standard reflects the aging of the population between 1970 and 2000 and gives more weight to older age groups. Consequently, the introduction of the new standard increased overall incidence rates by approximately 20% (**Table 1.4**). The effect of the new standard on the rate of a particular cancer type, however, depends on whether the cancer predominantly affects younger or older individuals. The change in standard increased rates by as much as 25% for cancer sites that increase with age such as lung and colon cancers, whereas they decreased by as much as 11% for childhood cancers such as acute lymphocytic leukemia. Age-adjusted rates from published reports using the 1970 standard should not be compared with rates based on the 2000 standard. The new standard does not affect either the age-specific rates or the total number of cases or deaths, but only the age-standardized rate.

Table 1.4
Effect of a change in standard population on age-adjusted incidence rates, United States, 2000

	Incidence rates		
	Standard population		
Cancer type	1970	2000	% change
All sites	397.8	472.9	18.9
Colon and rectum	42.4	53.1	25.1
Lung and bronchus	52.6	62.3	18.6
Female breast	115.2	135.1	17.2
Prostate	151.2	176.9	16.9
Acute lymphocytic leukemia	1.5	1.4	–11.0

3. Demographic Factors and Geographic Location

3.1. Age

Incidence rates exponentially increase with age for most cancers because of cumulative exposures to carcinogenic agents such as tobacco, infectious organisms, chemicals, and internal factors such as inherited mutations, hormones, and immune conditions. **Figure 1.2** (left) depicts the age-related increase in the incidence rate from all cancer combined in men and women. However, the incidence rate for age 0 to 4 is twice that for ages 5 to 9 and 10 to 14, and the rates of cancer diagnosis (if not true incidence) decrease after age 84. Specific cancers that contribute to higher incidence rates at ages 0 to 4, compared with 5 to 9 or 10 to 14 years, include acute lymphocytic leukemia, neuroblastoma, and retinoblastoma. The decrease in age 85 and older largely reflects

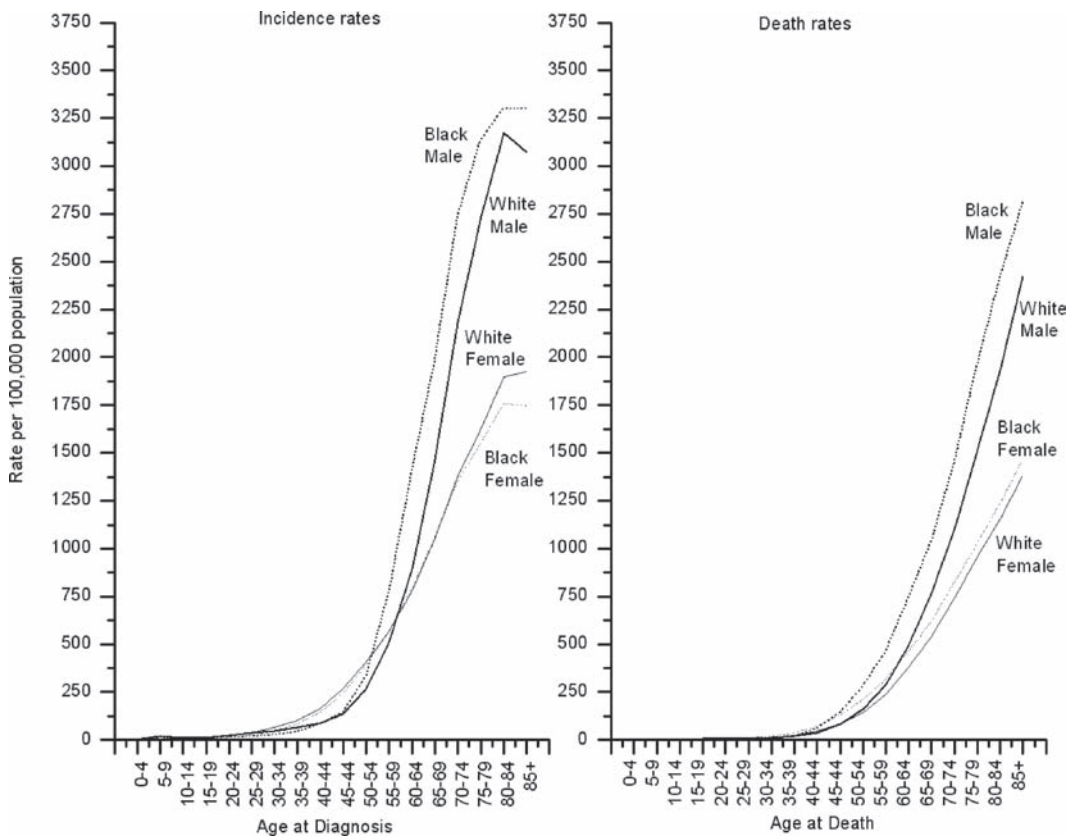


Fig. 1.2. Age- and sex-specific incidence and death rates from all cancers combined, United States, 2003. Incidence rates from SEER Program (http://www.seer.cancer.gov) SEER*Stat Database, 1973–2003. National Cancer Institute, 2006. Death rates from National Center for Health Statistics, Centers for Disease Control and Prevention, 2006.

underdiagnosis. Similar to the age-specific incidence patterns, age-specific mortality rates increase with age (**Fig. 1.2**, right). However, unlike the incidence patterns, the death rate at age 0 to 4 is not higher than rates at ages 5 to 9 and 10 to 14, and rates continue to increase through age 85 and older.

The median age at diagnosis for all cancers combined is 67 (**Fig. 1.3**). For individual cancer sites, it ranges from 34 years for testicular cancer to 73 years for gallbladder and urinary bladder cancer. The median age at death from all cancers combined is 73, with the median age for specific cancer sites ranging from 40 years for testicular cancer to 80 years for prostate cancer (**Fig. 1.4**). Although there are little differences between median age at diagnosis and median age at death for rapidly fatal cancers such as lung, pancreas, and liver, there are substantial differences between median age at diagnosis and median age at death for screenable cancers such as female breast cancer and for the most successfully treatable cancers such as leukemia.

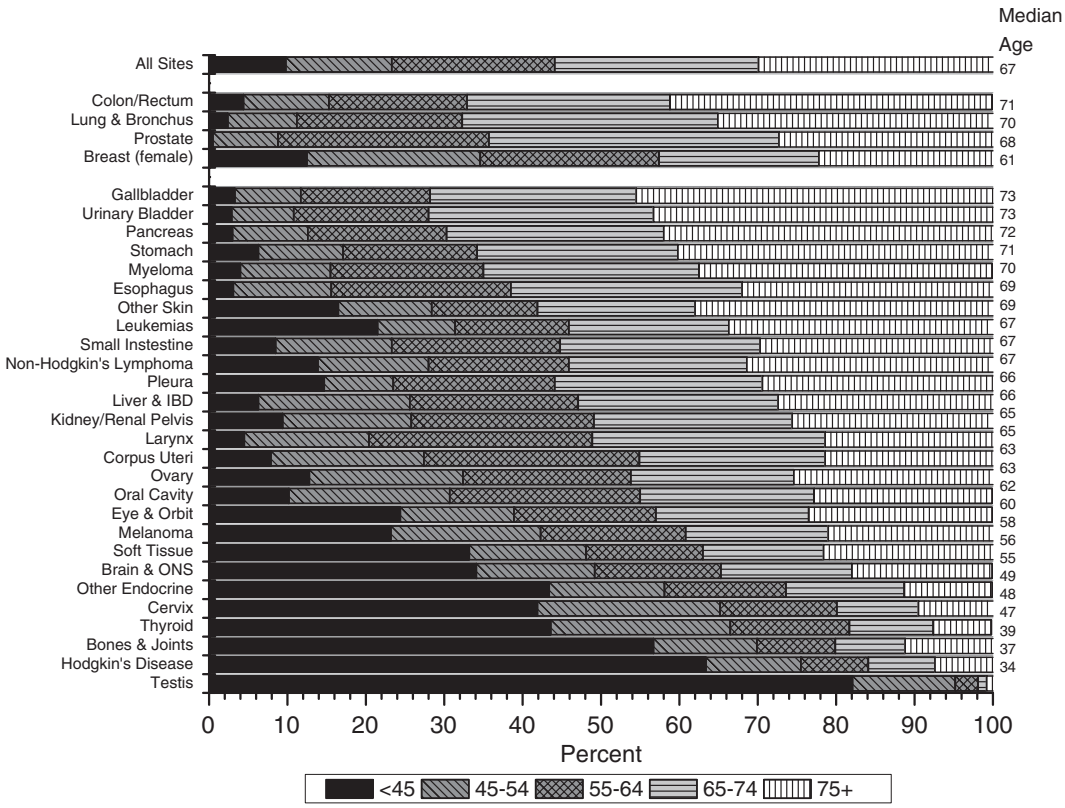


Fig. 1.3. Age distribution and median age at diagnosis for the incidence of major cancers, 2000–2003. Percentages may not add to 100 due to rounding. Data from SEER Program (http://www.seer.cancer.gov) SEER*Stat Database, 1973–2003. National Cancer Institute, 2006.

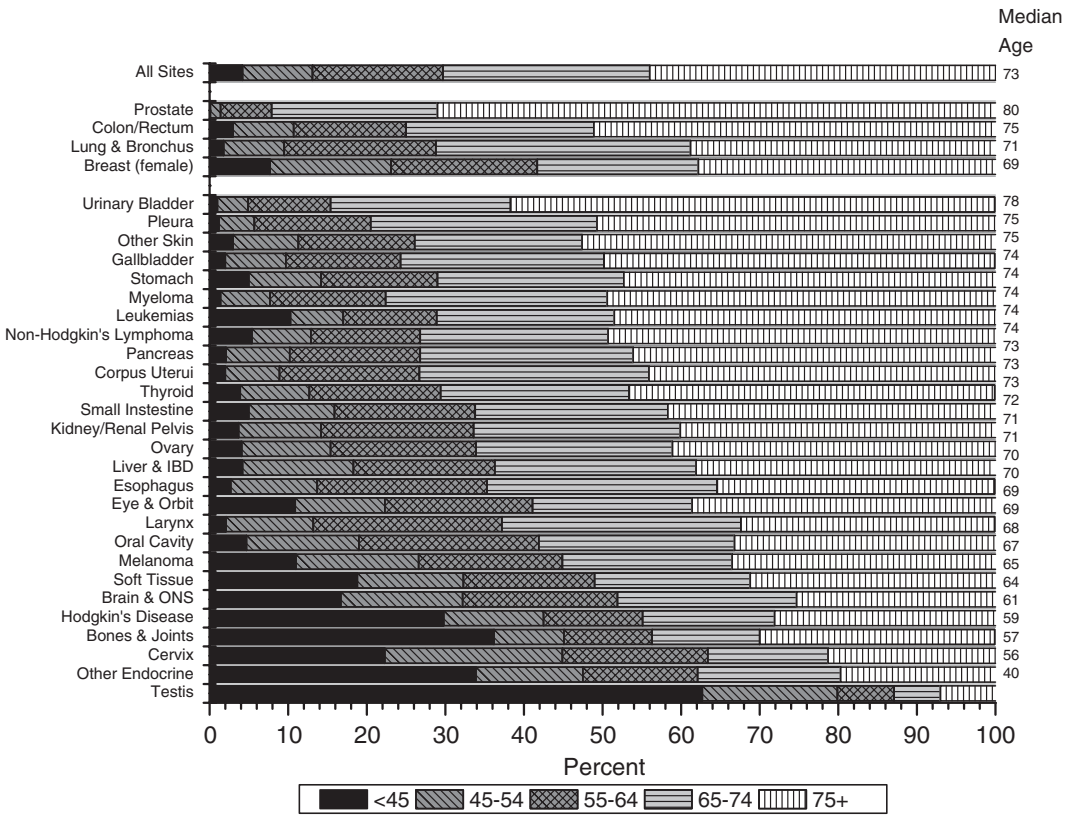


Fig. 1.4. Age distribution and median age at death for mortality due to major cancers, 2000–2003. Percentages may not add to 100 due to rounding. Data from National Center for Health Statistics, Centers for Disease Control and Prevention, 2006.

3.2. Sex

The incidence rates of most cancers that affect both men and women are higher in men than women (**Table 1.5**). The most extreme example is cancer of the larynx, for which the incidence rate is nearly five times as high in males as in females. The few exceptions in which cancers that affect both sexes are more common in women than men are breast, thyroid, and gallbladder. For breast cancer, the incidence rate is >100 times higher in women than men.

3.3. Socioeconomic Status

SES is inversely correlated with death rates from cancer and many other causes because it is strongly associated with high risk factors such as smoking, drinking, lack of health insurance, and low screening rates *(14–16)*. **Figure 1.5** shows the educational gradient in death rates from all cancers combined in ages 25–64-year-old non-Hispanic white men and women in 2001. The death rate in persons with ≤11 years of education compared with those with ≥16 years of education was 3 times higher in men and 2 times higher in women. In contrast, higher SES is positively associated with the incidence of some screening-related cancers. For

Table 1.5
Average annual age-standardized cancer incidence rates[a] in men compared with women, United States, 2000–2003

	Male rate	Female rate	Rate ratio (male vs. female)
All sites	558.13	412.02	1.35
Oral cavity and pharynx	15.68	6.12	2.56
Esophagus	7.76	1.99	3.90
Stomach	11.47	5.56	2.06
Small intestine	2.13	1.47	1.45
Colon and rectum	61.75	45.27	1.36
Anus, anal canal, and anorectum	1.30	1.60	0.81
Liver and intrahepatic bile duct	9.27	3.35	2.77
Gallbladder	0.83	1.46	0.57
Pancreas	12.79	10.01	1.28
Larynx	6.74	1.38	4.88
Lung and bronchus	82.15	52.28	1.57
Bones and Joints	1.06	0.77	1.38
Soft tissue including heart	3.65	2.59	1.41
Melanoma of the skin	23.17	14.73	1.57
Breast	1.25	129.14	0.01
Urinary bladder	37.03	9.29	3.99
Kidney and renal pelvis	17.48	8.68	2.01
Eye and orbit	0.99	0.65	1.52
Brain and other nervous system	7.62	5.38	1.42
Thyroid	4.22	12.06	0.35
Hodgkin lymphoma	3.04	2.35	1.29
Non-Hodgkin lymphoma	22.96	16.06	1.43
Myeloma	6.86	4.48	1.53
Leukemia	15.86	9.44	1.68

[a]Rates are per 100,000 and age-adjusted to the 2000 U.S. standard population.

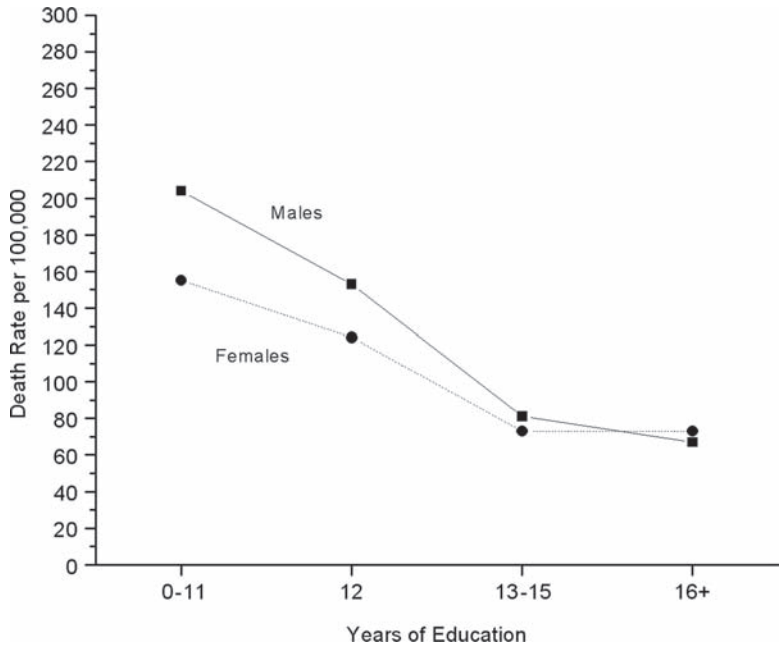


Fig. 1.5. Age standardized death rates for all cancers combined in non-Hispanic whites aged 25–64 by education level and sex, United States, 2001 (*Am J Prev Med* 34, 1–8, 2008).

example, the incidence of female breast and prostate cancers are higher in high and middle SES groups than in the poor, simply because of greater detection *(15,17)*.

3.4. Race and Ethnicity

Large differences in cancer incidence and death rates are seen by race and ethnicity in the United States and elsewhere *(18)*, although, in general, these differences reflect social, economic, and cultural factors rather than differences in inherited genetic susceptibility. Migrant studies have helped to differentiate whether variations in cancer rates across countries and among racial and ethnic groups are due to environmental factors or inherited genetic factors *(19)*. **Table 1.6** shows the difference in death rates from selected cancers comparing Japanese in Japan with first-(Issei) and second-(Nisei) generation Japanese in California, and with native California whites *(20)*. Mortality from cancers of the stomach and liver was much higher among Japanese in Japan than in California whites between 1950 and 1960. The opposite was true for cancer of the colon. However, among first generation Japanese men who migrated to California, the risk of dying from stomach or liver cancer was substantially lower than in Japan, although still higher than that of California whites. The risk among Japanese migrants and whites became more similar by the second generation. In contrast, the risk of death from colon cancer increased rapidly after migration from Japan to California,

Table 1.6
Stomach, colon, and liver cancer risks in Japanese in Japan and in California, and California whites age 45–64, between 1956 and 1962

		Japanese		
Cancer site	Whites California	2nd generation in California	1st generation in California	Japan
Stomach	1.0	2.8	3.8	8.4
Colon	1.0	0.9	0.4	0.2
Liver	1.0	2.2	2.7	4.1

Source: *Cancer* **18,** 656–64, 1965.

approximately doubling in the first-generation Japanese who moved to California and approaching the rates of California white men by the second generation. Currently, colorectal cancer death rates among Japanese in both the United States and Japan are higher than rates in whites, presumably reflecting changing dietary and physical activity patterns in Asia as well as among migrants to the United States *(21,22)*.

Table 1.7 illustrates the large difference in the incidence and death rates between black and white Americans in the time period 1999–2003. The incidence of all cancers combined was 25% higher in black than white men; the mortality rate was 43% higher. Similarly, the death rate from all cancers combined was nearly 20% higher in black than white women, despite lower incidence rates. These differences reflect a combination of disparities that affect prevention, early detection, and treatment. The extent to which inherited differences in cancer susceptibility contribute to the observed differences is not yet clear, but it is likely to be very small. In contrast, other major racial ethnic groups in the United States have lower incidence and mortality for all sites combined and for the four most common cancer sites (lung and bronchus, colon and rectum, prostate, and female breast) compared with the rates in whites and blacks. For certain cancer sites, however, the incidence in Hispanic and Asian migrants remains higher than that in whites. This is particularly true for cancers of the stomach, liver, cervix uteri, and intrahepatic bile duct. All of these cancers are affected by specific infectious agents that are more prevalent in the countries of origin than in the United States.

3.5. Geographic Location

Variability in cancer occurrence by place, together with migrant studies and surveillance data on cancer trends, has stimulated important hypotheses about the etiology and potential

Table 1.7
Age-standardized incidence and death rates[a] for selected cancer sites by race and ethnicity, United States, 1999–2003

Incidence	White	African American	Asian/Pacific Islander	American Indian/ Alaskan Native[b]	Hispanic-Latino[c]
All sites					
Males	555.0	639.8	385.5	359.9	444.1
Females	421.1	383.8	303.3	305.0	327.2
Breast (female)	130.8	111.5	91.2	74.4	92.6
Colon and rectum					
Males	63.7	70.2	52.6	52.7	52.4
Females	45.9	53.5	38.0	41.9	37.3
Lung and bronchus					
Males	88.8	110.6	56.6	55.5	52.7
Females	56.2	50.3	28.7	33.8	26.7
Prostate	156.0	243.0	104.2	70.7	141.1
Stomach					
Males	9.7	17.4	20.0	21.6	16.1
Females	4.4	9.0	11.4	12.3	9.1
Liver and bile duct					
Males	7.2	11.1	22.1	14.5	14.8
Females	2.7	3.6	8.3	6.5	5.8
Uterine cervix	8.6	13.0	9.3	7.2	14.7
Mortality	**White**	**African American**	**Asian/Pacific Islander**	**American Indian/ Alaskan Native**	**Hispanic-Latino ‡**
All sites					
Males	239.2	331.0	144.9	153.4	166.4
Females	163.4	192.4	98.8	111.6	108.8
Breast (female)	25.4	34.4	12.6	13.8	16.3
Colon and rectum					
Males	23.7	33.6	15.3	15.9	17.5
Females	16.4	23.7	10.5	11.1	11.4
Lung and bronchus					
Males	73.8	98.4	38.8	42.9	32.7

Table 1.7 (continued)

Mortality	White	African American	Asian/Pacific Islander	American Indian/ Alaskan Native[b]	Hispanic-Latino[c]
Females	42.0	39.8	18.8	27.0	14.7
Prostate	26.7	65.1	11.8	18.0	22.1
Stomach					
Males	5.4	12.4	11.0	7.1	9.2
Females	2.7	6.0	6.7	3.7	5.2
Liver and bile duct					
Males	6.3	9.6	15.5	7.8	10.7
Females	2.8	3.8	6.7	4.0	5.0
Uterine cervix	2.4	5.1	2.5	2.6	3.4

[a]Per 100,000, age-adjusted to 2000 U.S. standard population.[b]Incidence rate are for diagnosis years 1999–2002.[c]Hispanic-Latinos are not mutually exclusive from whites, blacks, Asian/Pacific Islanders, and American Indian/Alaska Natives.Source: Incidence (except American Indian and Alaska Native): *Cancer* **107,** 1643–1658, 2006. Incidence (American Indian and Alaska Native) and mortality: Reis, L. A. G., Harkins, D., Krapcho, M., et al. (eds.) SEER Cancer Statistics Review, 1975–2003, National Cancer Institute, based on November 2005 SEER data submission, posted to the SEER website, 2006.

preventability of many cancers *(1,23)*. **Table 1.8** shows death rates from certain cancers in the United States and selected other countries. The age-adjusted rates were calculated using the world standard population for 1960 rather than the U.S. standard population. The large variations in death rate of the specific cancers across countries may reflect a combination of differences in the prevalence of underlying risk factors, differences in host susceptibility and/or variations in detection, completeness of reporting, treatment, and classification of disease. For example, the especially high lung cancer rates in eastern European countries, such as Hungary and the Czech Republic, reflect very high prevalence of long-term cigarette smoking. The cervical cancer death rate is highest in Mexico and Central and South America because of lack of screening and early treatment for cervical cancer. In contrast, female breast cancer death rates are higher in Scandinavian countries and Western Europe, in part due to increased underlying risk factors, such as late age at childbearing. The highest female breast cancer death rates in Israel may reflect greater genetic susceptibility.

High-risk areas for specific cancers may or may not be well characterized by official administrative boundaries, such as county, state, or national borders. For example, in the United States,

Table 1.8
International variations in cancer death rates for selected cancer sites, GLOBOCAN 2002

	Lung		Colon/rectum		Stomach		Breast	Cervix uteri	Prostate
Country	Male	Female	Male	Female	Male	Female	Female	Female	Male
USA	48.7	26.8	15.2	11.6	4.0	2.2	19.0	2.3	15.8
Chile	21.0	7.6	7.7	7.8	32.5	13.2	13.1	10.9	20.8
China	36.7	16.3	7.9	5.3	32.7	15.1	5.5	3.8	1.0
Czech Republic	61.8	12.8	34.0	18.0	12.1	6.4	20.0	5.5	17.2
Denmark	45.2	27.8	23.3	19.2	5.5	3.3	27.8	5.0	22.6
France	47.5	8.0	18.2	11.8	7.0	3.1	21.5	3.1	18.2
Germany	42.4	10.8	19.9	15.7	10.4	6.4	21.6	3.8	15.8
Hungary	83.9	22.3	35.6	21.2	18.2	8.5	24.6	6.7	18.4
Israel	26.9	8.6	18.8	14.6	9.8	4.7	24.0	2.3	13.5
Japan	32.4	9.6	17.3	11.1	28.7	12.7	8.3	2.8	5.7
Mexico	16.6	6.6	4.5	4.1	9.9	7.2	10.5	14.1	14.9
Norway	32.7	13.5	20.1	16.8	9.4	5.0	17.9	3.5	28.4
Russian Federation	36.0	6.2	18.9	13.6	31.8	13.5	18.0	6.5	8.2
Spain	49.2	4.7	18.5	11.3	11.4	5.4	15.9	2.2	14.9
United Kingdom	42.9	21.1	17.5	12.4	8.7	4.0	24.3	3.1	17.9
Venezuela	18.1	10.2	6.4	6.7	14.5	9.3	13.4	16.8	19.8

the area with the highest death rates from cervical cancer spans much of Appalachia *(24)* where women historically lacked access to regular pap testing or treatment. This observation motivated the U.S. Congress to create the National Breast and Cervical Cancer Early Detection Program (NBCCEDP) to improve access to breast and cervical cancer screening and diagnostic services for low-income women *(25)*. Although this program currently serves only 13% of eligible women because of insufficient funding for NBCCEDP, it illustrates how a national or multinational approach may be needed for public health problems that extend across state or national borders.

Geographic information system (GIS) analysis is a relatively new tool for describing cancer patterns by place of residence, by

using the address at diagnosis or death. It is a powerful automated system for the capture, storage, retrieval, analysis, and display of spatial data. Examples of GIS application in cancer epidemiology include identification of geographic areas with a high proportion of distant stage breast cancers in New Jersey *(26)* and finding an inverse relationship between travel distance to the nearest radiation therapy facility and receipt of radiotherapy after breast-conserving surgery in New Mexico *(27)*.

3.6. Temporal Trends

Cancer incidence and mortality rates have substantially changed over the years (**Fig. 1.6**; *see* **ref**. *28)*. The increase in the incidence rate for all cancers combined among men from 1975 through 1988 reflects the rise in lung cancer from smoking and the diagnosis of prostate cancer from transurethral resection *(28,29)*. In contrast, the sharp increase and subsequent decrease in the

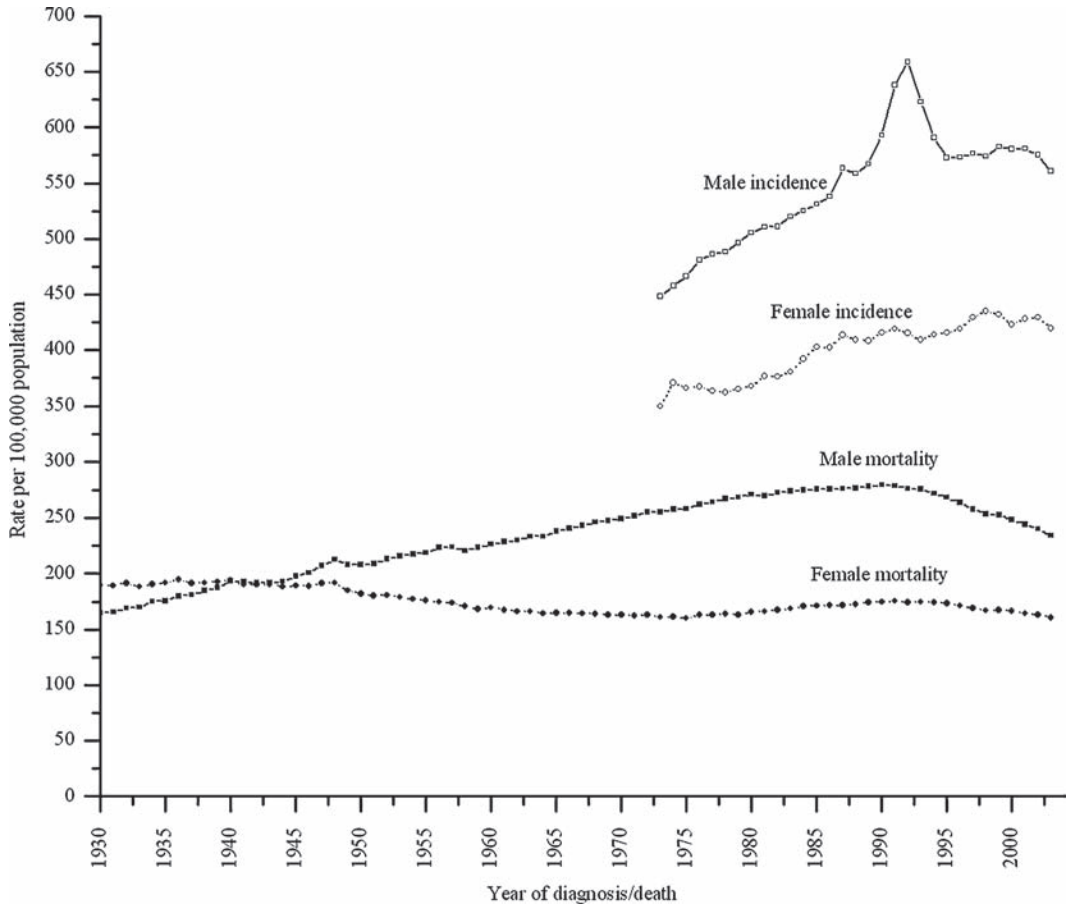


Fig. 1.6. Trends in all cancers, combined incidence and mortality, United States. Rates are age adjusted to the 2000 U.S. standard population. Incidence rates from SEER Program (http://www.seer.cancer.gov) SEER*Stat Database, 1973–2003. National Cancer Institute, 2007. Death rates from National Center for Health Statistics, Centers for Disease Control and Prevention, 2006.

incidence of all cancers combined from 1988 through 1992 reflect the introduction of PSA testing followed by the saturation and leveling off of PSA testing *(30)*. The overall increase in incidence rates among women (**Fig. 1.6**), although slower during the most recent time period, predominantly reflects trends in lung and breast cancer incidence, which in turn result from the increased consumption of cigarettes in women born in the 1930s *(29)* and increased use of mammography since the mid-1980s *(31)*, respectively. Female breast cancer incidence rates are declining since 2000, likely due to reduction in use of mammography and hormone replacement therapy *(32,33)*.

Among men, the long-term trend in overall cancer death rates (**Fig. 1.6**) is largely determined by trends in tobacco related cancers, particularly lung cancer (**Fig. 1.7**, left; see **ref**. *34)*. Between 1930 and 1990, lung cancer death rates (per 100,000) rose from 4.3 to 90.4, representing a 21-fold increase. Thereafter,

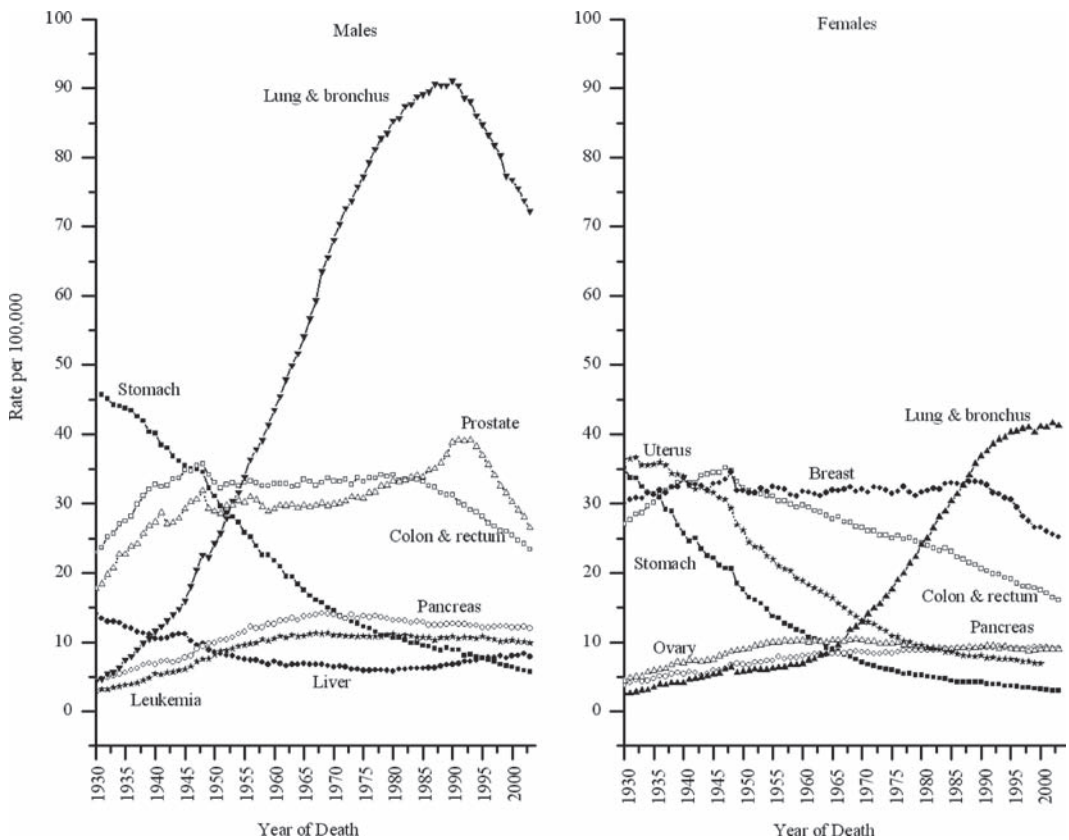


Fig. 1.7. Trends in annual age-adjusted cancer death rates among males and females for selected cancer types, United States, 1930 to 2003. Rates are age adjusted to the 2000 U.S. standard population. Data from National Center for Health Statistics, Centers for Disease Control and Prevention, 2006. Due to changes in ICD coding, numerator information has changed over time. Rates for cancers of the liver, lung and bronchus, colon and rectum, and ovary are affected by these coding changes. Uterus cancer death rates are for uterine cervix and uterine corpus combined.

lung cancer death rates decreased to 70.3 per 100,000 in 2004. These trends reflect historical smoking patterns. Cigarette consumption peaked during the mid-twentieth century and gradually decreased after the surgeon general's report in 1964 *(35)*. In addition to lung cancer, cigarette smoking has been linked to multiple other cancers types, including oral cavity, pharynx, larynx, esophagus, stomach, pancreas, urinary bladder, kidney, liver, cervix, and myeloid leukemia.

Improved treatments, increased screening, and changes in dietary habits also may have contributed to the reduction in death rates from all cancers combined since the early 1990s. Colorectal and prostate cancers are among the sites for which trends are shown in **Fig. 1.7** that may be affected by a combination of improved treatment, screening, or both.

Among women, the overall cancer death rate decreased from the 1930s through early 1970s, increased until the early 1990s, and decreased thereafter (**Fig. 1.6**). The long-term decline before the 1970s reflects decreases in deaths from cancer of the stomach, cervix-uterus, and colon and rectum (**Fig. 1.7**, right). The dramatic decrease in stomach cancer mortality, which has occurred in most industrialized countries, is thought to result from improvements in food preservation and reduced prevalence of *Helicobacter pylori* infection *(36)*. Among women, the historic decreases in mortality from stomach and cervix-uterus cancers were offset by the increases in lung cancer mortality after 1960 (**Fig. 1.7**, right, *see* **ref**. *29)*. Decreases in the overall cancer death rates since the early 1990s reflect reduction in death rates from colorectal and breast cancers due to improved treatment and increased screening rates and plateauing of trends in lung mortality rates. It is interesting to note that death rates from certain cancers were higher in women than in men early in the twentieth century, presumably because of higher death rates from female reproductive cancers.

4. Issues in Interpretation of Temporal Trends

Decreases in incidence rates not only reflect the impact of primary preventive strategies, such as reduction in cigarette smoking, but also may result from increased screening for cancers with known precursors (polyps for colorectal cancer and superficial cellular abnormalities for cervical cancer). An increase in the incidence rates may signal either increased exposure to risk factors or increased detection due to introduction of screening. Reductions in mortality are easier to interpret and less susceptible to artifact than are trends in incidence and survival. Such decreases generally reflect the effectiveness of cancer prevention, early detection, and

treatment improvements over past decades. Trends in incidence rates also can be influenced by methodologic factors, including delay in reporting of cancer cases to registries, revision of population estimates, and changes in diagnosis and disease classification; of these, the latter two factors also may affect mortality trends.

Incidence rates and trends are known to be affected by delays in reporting *(37)*. Population-based registries usually calculate and report rates 3 years after the year of diagnosis, to allow time for data collection and coding. However, cancer types that are commonly diagnosed in outpatient settings, such as leukemia, prostate cancer, and melanoma, may be reported to registries even later. A method has been developed to adjust for delayed reporting in analyzing cancer incidence trends from the SEER registry database *(37)*. Delayed reporting may cause the incidence rates for melanoma and prostate cancer to be underestimated by as much as 14% in the most recent 1 to 2 years *(37)*. It should be noted that the impact of delays in reporting in SEER registry data is probably less than the impact for other population-based registries, because they are long-standing registries of very high quality.

The introduction of new diagnostic or screening techniques also can influence trends in cancer incidence rates. Typically, the introduction of a new screening technique leads to a rise in incidence rate for a brief period, with subsequent decline and return to the baseline; this outcome is because existing cases are diagnosed earlier than would have been the case in the absence of screening. For example, the rapid increase and subsequent decrease in prostate cancer incidence rates between 1988 and 1992 largely reflects the broad dissemination and use of PSA testing *(30)* rather than a real increase in disease occurrence. The corresponding rise in prostate cancer mortality, which was unexpected, is thought to result from attribution bias by physicians in recording cause of death *(38)*.

Change in disease classification also may influence interpretation of incidence and mortality. The World Health Organization updates international statistical classification of diseases approximately every 10 years to stay abreast of medical advances. The introduction of ICD-10 beginning with deaths occurring in 1999 increased the proportion of deaths attributed to cancer by 0.7% relative to ICD-9 *(39)*. This change disrupted the downward trend in mortality that began in 1992, with the decline seeming to end in 1998 *(34)*. When the change in ICD-10 was accounted for, however, the decrease in rate continued from 1992 to 2000 *(40)*.

4.1. Understanding Temporal Trends

As discussed above, long-term temporal trends in cancer incidence rates can be influenced by several factors, the most prominent of which are changes in underlying risk factors and introduction

of new screening techniques. Mortality is influenced by these factors as well as by the introduction of more effective treatments.

In describing temporal trends, epidemiologists often refer to time period and cohort effects. A period effect results from medical advances (e.g., the introduction of a new screening technique or improved treatment) or changes in disease classification that increase or decrease the incidence rate across all age groups in the same calendar period. A striking contemporary example of a period effect is the sharp increase and subsequent decrease in prostate cancer incidence rates in almost all age groups between the late 1980s and early 1990s in the United States (**Fig. 1.8**, left), reflecting the introduction and broad use of PSA testing in the late 1980s (30).

In contrast, a birth cohort effect typically results from the introduction (or increased prevalence) of a risk factor that becomes established at a young age in people born during the same time period. In birth cohort patterns, the disease consequences from

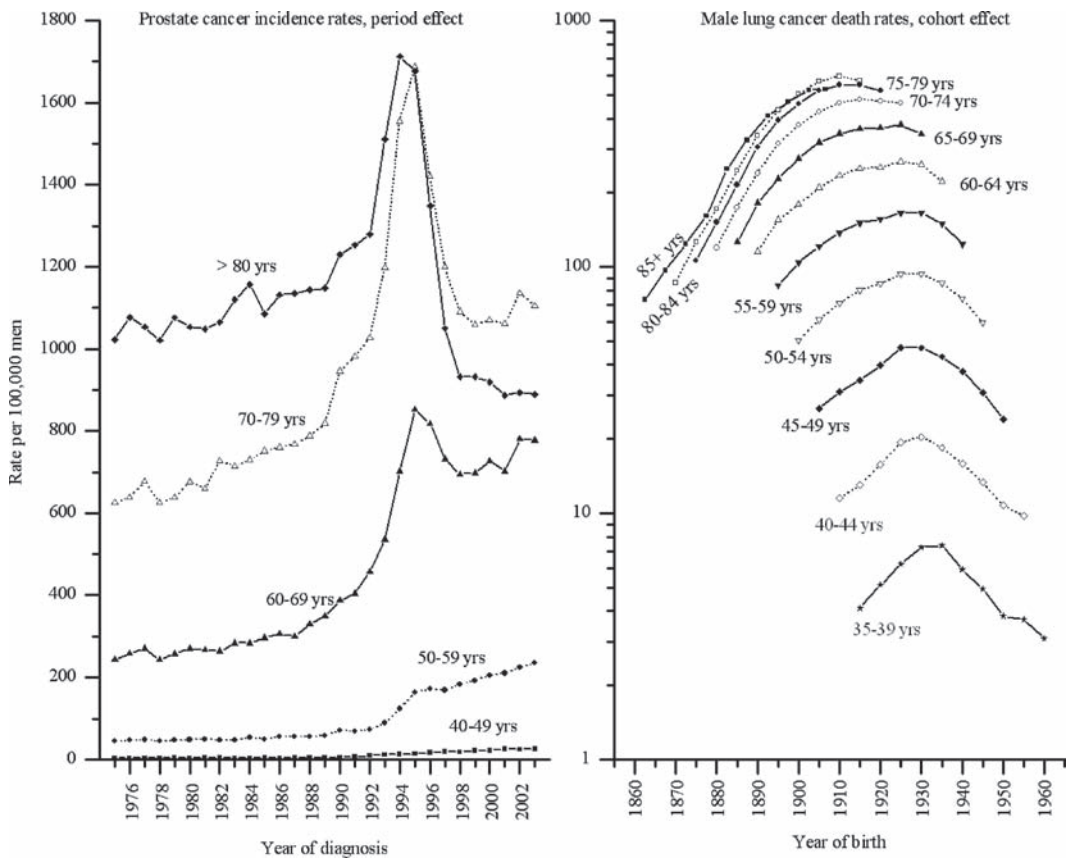


Fig. 1.8. Trends in age-specific rates showing period and birth cohort effects. Incidence rates from SEER Program (http://www.seer.cancer.gov) SEER*Stat Database, 1973–2003. National Cancer Institute, 2006. Death rates from National Center for Health Statistics, Centers for Disease Control and Prevention, 2006.

these exposures become manifest later in life as each birth cohort ages. Birth cohort effects are often identified by plotting the age-specific rates against year of birth, so that birth cohorts with higher exposure have higher incidence or death rates at a given age. For example, the birth cohort pattern in lung cancer mortality in men is depicted in **Fig. 1.8** (right). Age-specific male lung cancer death rates peaked in cohorts born around the late 1920s and early 1930s among all age groups. Cigarette smoking prevalence also peaked in cohorts born around this period. Smoking rates were highest among males who passed through adolescence and young adulthood at a time when cigarettes were plentiful and aggressive marketing was unopposed by health concerns. Smoking prevalence declined in the subsequent birth cohorts because of growing publicity about the adverse effects of smoking on health *(41)*. As is the case for cigarette smoking and lung cancer, many temporal trends combine time period and cohort effects. These can be analyzed using an age-period-cohort model to partition the contribution of period and cohort effects. This method is described elsewhere *(42,43)*.

References

1. Doll, R., and Peto, R. (1981) The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today *J Natl Cancer Inst* **66,** 1191–308.
2. Percy, C, Van Holten, V., and Muir, C. (1990) *International Classification of Diseases for Oncology* (ICD-O), 2nd ed, World Health Organization, Geneva, Switzerland.
3. U.S. Cancer Statistics Working Group (2006) *United States Cancer Statistics: 2003 Incidence and Mortality*, U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute, Atlanta, GA.
4. World Health Organization (1992) *International Statistical Classification of Diseases and Related Health Problems*, World Health Organization, Geneva, Switzerland.
5. Kircher, T., Nelson, J., and Burdo, H. (1985) The autopsy as a measure of accuracy of the death certificate *N Engl J Med* **313,** 1263–9.
6. Ferlay, J., Bray, F., Pisani, P., and Parkin, D. M. (2004) IARC Press, Lyon, France.
7. Parkin, D. M., Bray, F., Ferlay, J., and Pisani, P. (2001) Estimating the world cancer burden: Globocan 2000. *Int J Cancer* **94,** 153–6.
8. Tiwari, R. C., Ghosh, K., Jemal, A., Hachey, M., Ward, E., Thun, M. J., and Feuer, E. J. (2004) Derivation and validation of a new method of predicting U.S. and state-level cancer mortality counts for the current calendar year *CA Cancer J Clin* **54,** 8–29.
9. Wingo, P. A., Landis, S., and Parker, S. (1998) Using cancer registry and vital statistics to estimate the number of new cases cancer cases and deaths in the US for the upcoming year. *J Regist. Manage.* **25,** 43–51.
10. Edwards, B. K., Howe, H. L., Ries, L. A., Thun, M. J., Rosenberg, H. M., Yancik, R., Wingo, P. A., Jemal, A., and Feigal, E. G. (2002) Annual report to the nation on the status of cancer, 1973–1999, featuring implications of age and aging on U.S. cancer burden *Cancer* **94,** 2766–92.
11. Ries, L. A. G., Melbert, D., Krapcho, M., Mariotto, A., Miller, B. A., Feuer, E. J., Clegg, L., Horner, M. J., Howlader, N., Eisner, M. P., Reichman, M., and Edwards, B. K. (eds.) (2007) *SEER Cancer Statistics Review, 1975–2004*, National Cancer Institute, National Cancer Institute, Bethesda, MD.
12. Clegg, L. X., Li, F. P., Hankey, B. F., Chu, K., and Edwards, B. K. (2002) Cancer survival among US whites and minorities: a SEER (Surveillance, Epidemiology, and End Results) Program population-based study. *Arch Intern Med* **162,** 1985–93.
13. Anderson, R. N., and Rosenberg, H. M. (1998) Age standardization of death rates: implementation of the year 2000 standard. *Natl Vital Stat Rep* **47,** 1–16, 20.

14. Ward, E., Jemal, A., Cokkinides, V., Singh, G. K., Cardinez, C., Ghafoor, A., and Thun, M. (2004) Cancer disparities by race/ethnicity and socioeconomic status. *CA Cancer J Clin* **54,** 78–93.
15. Singh, G. K., Miller, B. A., Hankey, B. F., and Edwards, B. K. (2003), pp. 137, National Cancer Institute, Bethesda, MD.
16. Howard, G., Anderson, R. T., Russell, G., Howard, V. J., and Burke, G. L. (2000) Race, socioeconomic status, and cause-specific mortality. *Ann Epidemiol* **10,** 214–23.
17. Kitagawa, E., and Hauser, P. (1973) *Differential Mortality in the United States: A Study in Socioeconomic Epidemiology*, Harvard University Press, Cambridge, MA.
18. Kogevinas, M., Pearce, N., Susser, M., and Boffetta, P. (1997) *Social Inequalities and Cancer*, International Agency for Research on Cancer, Lyon, France.
19. Kolonel, L., and Wilkens, L. (2006) Migrant studies in *Cancer Epidemiology and Prevention, 3rd ed*., Oxford University Press, New York, pp. 189–201.
20. Buell, P., and Dunn, J. (1965) Cancer mortality among Japanese Issei and Nisei of California. *Cancer* **18,** 656–64.
21. Bernstein, L., and Wu, A. H. (1998) Colon and Rectal Cancer among Asian Americans and Pacific Islanders. *Asian Am Pac Isl J Health* **6,** 184–200.
22. Stewart, B. W., and Kleihues, P. (eds.) (2003) *World Cancer Report*, IARC Press, Lyon, France.
23. Doll, R. (1998) Epidemiological evidence of the effects of behaviour and the environment on the risk of human cancer. *Recent Results Cancer Res* **154,** 3–21.
24. Devesa, S. S., Grauman, D. J., Blot, W. J., Pennello, G. A., Hoover, R. N., and Fraumeni, J. F. (1999) *Atlas of Cancer Mortality in the United States 1950–94*, National Institute of Health, Bethesda, MD.
25. Centers for Disease Control and Prevention (2003) http://www.cdc.gov/cancer/nbccedp/about.htm.
26. Roche, L. M., Skinner, R., and Weinstein, R. B. (2002) Use of a geographic information system to identify and characterize areas with high proportions of distant stage breast cancer. *J Public Health Manag Pract* **8,** 26–32.
27. Athas, W. F., Adams-Cameron, M., Hunt, W. C., Amir-Fazli, A., and Key, C. R. (2000) Travel distance to radiation therapy and receipt of radiotherapy following breast-conserving surgery. *J Natl Cancer Inst* **92,** 269–71.
28. Potosky, A. L., Kessler, L., Gridley, G., Brown, C. C., and Horm, J. W. (1990) Rise in prostatic cancer incidence associated with increased use of transurethral resection. *J Natl Cancer Inst* **82,** 1624–8.
29. Thun, M. J., Henley, S. J., and Calle, E. E. (2002) Tobacco use and cancer: an epidemiologic perspective for geneticists *Oncogene* **21,** 7307–25.
30. Potosky, A. L., Miller, B. A., Albertsen, P. C., and Kramer, B. S. (1995) The role of increasing detection in the rising incidence of prostate cancer *Jama* **273,** 548–52.
31. Smigal, C., Jemal, A., Ward, E., Cokkinides, V., Smith, R., Howe, H. L., and Thun, M. (2006) Trends in breast cancer by race and ethnicity: update 2006 *CA Cancer J Clin* **56,** 168–83.
32. Jemal, A., Ward, E., and Thun, M. J. (2007) Recent trends in breast cancer incidence rates by age and tumor characteristics among U.S. women. *Breast Cancer Res* **9,** R28.
33. Ravdin, P. M., Cronin, K. A., Howlader, N., Berg, C. D., Chlebowski, R. T., Feuer, E. J., Edwards, B. K., and Berry, D. A. (2007) The decrease in breast-cancer incidence in 2003 in the United States. *N Engl J Med* **356,** 1670–4.
34. Howe, H. L., Wu, X., Ries, L. A., Cokkinides, V., Ahmed, F., Jemal, A., Miller, B., Williams, M., Ward, E., Wingo, P. A., Ramirez, A., and Edwards, B. K. (2006) Annual report to the nation on the status of cancer, 1975–2003, featuring cancer among U.S. Hispanic/Latino populations. *Cancer* **107,** 1711–42.
35. U.S. Department of Health and Human Services (2004) *A Surgeon General's Report on the Health Consequences of Smoking*, U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Office of Smoking and Health, Atlanta, GA.
36. Parkin, D. M. (2006) The global health burden of infection-associated cancers in the year 2002. *Int J Cancer* **118,** 3030–44.
37. Clegg, L. X., Feuer, E. J., Midthune, D. N., Fay, M. P., and Hankey, B. F. (2002) Impact of reporting delay and reporting error on cancer incidence rates and trends. *J Natl Cancer Inst* **94,** 1537–45.
38. Feuer, E. J., Merrill, R. M., and Hankey, B. F. (1999) Cancer surveillance series: interpreting trends in prostate cancer II: cause of death misclassification and the recent rise and fall in prostate cancer mortality. *J Natl Cancer Inst* **91,** 1025–32.

39. Anderson, R. N., Minino, A. M., Hoyert, D. L., and Rosenberg, H. M. (2001) Comparability of cause of death between ICD-9 and ICD-10: preliminary estimates. *Natl Vital Stat Rep* **49,** 1–32.
40. Jemal, A., Ward, E., Anderson, R., and Thun, M. (2003) The influence of ICD-10 on U.S. cancer mortality trends. *J Natl Cancer Inst* **95,** 1727–8.
41. U.S. Public Health Service (1964) *Smoking and Health. Report of the Advisory Committee to the Surgeon General of the Public Health Service*, Department of Health, Education, and Welfare, Public Health Service, Center for Disease Control, Washington, DC.
42. Clayton, D., and Schifflers, E. (1987) Models for temporal variation in cancer rates. II: age-period-cohort models. *Stat Med* **6,** 469–81.
43. Clayton, D., and Schifflers, E. (1987) Models for temporal variation in cancer rates. I: age-period and age-cohort models. *Stat Med* **6,** 449–67.

Chapter 2

Cancer Registry Databases: An Overview of Techniques of Statistical Analysis and Impact on Cancer Epidemiology

Ananya Das

Summary

Cancer registries provide systematically collected information on cancer incidence, prevalence, mortality, and survival of different cancers. Aggregated and de-identified patient-level information on cancer is available for analysis from individual cancer registries, nationally from the Surveillance, Epidemiology, and End Results program, the Centers for Diseases Control and Prevention, the North American Association of Central Cancer Registries; and internationally from the International Association of Cancer Registries. Over the past few decades, the type and extent of cancer-related information captured by different cancer registries have been greatly expanded by linkage with other population-based information sources, such as the census data and the Centers for Medicare and Medicaid Services claims data. In addition, sophisticated statistical analytical techniques have been developed that have greatly expanded the traditional purview of cancer registries focused on descriptive epidemiology and disease quantification to a much broader analytical horizon ranging from study of cancer etiology; rare cancers in specific demographic groups; interaction of environmental and genetic factors in causation of cancer; impact of co-morbidities, race, geographic, socioeconomic, and provider-related factors on access, diagnosis, and treatment; outcomes and end results of cancer treatment; and cancer control initiatives to diverse areas of cancer care disparity, public health policy, public health education, and importantly, cost-effectiveness of cancer care. Thus, it is not surprising that cancer registries have increasingly become indispensable parts of local, national, and international cancer control programs, and it is certain that cancer registries will continue to be extraordinary resources of information for clinicians, researchers, scientists, policy makers, and the public in our fight against cancer.

Key words: Cancer, cancer registry, epidemiology, SEER, SEER-Medicare database.

1. Introduction

Cancer is a leading cause of death worldwide, and it accounted for more than a quarter of all deaths in the United States in 2003. One of the essential tools in the fight against cancer is systemically collected information on all aspects of cancer. Early attempts at

Mukesh Verma (ed.), *Methods in Molecular Biology, Cancer Epidemiology, vol. 471*
© 2009 Humana Press, a part of Springer Science + Business Media, Totowa, NJ
Book doi: 10.1007/978-1-59745-416-2

systemic collection of information on cancer date back to the early eighteenth century; the first American cancer registry was established in the 1920s, and the first national cancer registry was the Danish cancer registry established in 1942 *(1)*. Since then, there has been explosive growth in the number of population based cancer registries and, currently approximately 450 member registries contribute to the International Association Cancer Registries (IACR) representing approximately 21% of the world population *(2)*.

2. Cancer Registry Databases

Although the original purpose of the data collected by cancer registries was to generate statistics on incidence of cancer, over the past several decades both the type of data that is being collected and the analytical information based on available data that is being reported have greatly expanded to examine nearly every aspect of cancer epidemiology. These aspects include but are not limited to statistical analysis of cancer incidence, mortality, survival, and risk factors; hypothesis generation regarding cancer etiology; studying rare forms of cancer, in specific patient groups; monitoring programs for screening and surveillance, and treatment outcomes; and identifying disparities across population subgroups with respect to race, socioeconomic, and demographic variables.

Cancer registry databases provide a wealth of data for research in cancer epidemiology. Aggregated and de-identified patient-level information on cancer is available for analysis from individual cancer registries, nationally from the Surveillance, Epidemiology, and End Results (SEER) Program, the Centers for Diseases Control and Prevention (CDC), the North American Association of Central Cancer Registries (NAACCR); and internationally from the International Association of Cancer Registries (IACR).

One of the most important requirements of cancer surveillance data is that the registry data need to be standardized in terms of coding practices, case identification, and conversion of medical terminology to appropriate categories for enabling comparison across registries, countries, and even over time for studying trend. Currently, most population based cancer registries have standardized operating procedures for collection of surveillance data that is reasonably accurate, complete, valid and timely; however, there is still considerable variability in the quality of data available from different cancer registries *(3)*.

The type and extent of cancer-related information captured by different cancer registries is quite variable and limited by available resources. The essential components of any cancer registry data

are personal identification components (for linkage); demographic variables; and incidence date of cancer with most valid diagnosis, site and morphology of cancer, tumor behavior, and source of information. Recommended variables are follow-up data, date and status at last contact, stage and extent of disease at diagnosis, and initial treatment provided *(4,5)*.

Because cancer registries are gradually becoming an essential component of national cancer control programs, an increasing amount of resources is being committed to the cancer registries; and correspondingly, the capture of information on cancer patients has been going beyond the confines of traditional registry data set and incorporating details of clinical treatment and also useful socioeconomic information.

With the widespread availability and use of sophisticated computerized databases, it has become possible to link cancer registry information to other population-based information sources that not only provide population counts as cancer rate denominators but also provide patient-level data on co-morbidities, risk factors, treatment outcomes, and access to care; and linkage to census data at the neighborhood level (e.g., zip code, census track, block-group); and other geographical information systems have enabled researchers to answer complex issues such socioeconomic disparity in cancer care to and test etiologic hypotheses. For example, SEER uses the Population Estimates Program data of the U.S. Census Bureau and U.S. mortality data, collected and maintained by the National Center for Health Statistics, for population counts to be used as denominator for calculation of cancer incidence and mortality *(6)*. Similarly, linkage of SEER data to the Medicare database and to AIDS registries has enabled researchers to study cancer in Americans eligible for Medicare *(7)* and AIDS-related cancers *(8)*, respectively. Similarly, population based databases, such as, Behavioral Risk Factor Surveillance System of the CDC *(9)*, the National Health Interview Survey *(10)*, the National Hospital Discharge survey *(11)*, and the National Ambulatory Medical Care Surveys *(12)* are available for characterization of risk factors and cancer screening behaviors.

3. Cancer Registry Databases in the United States

In the United States, the National Cancer Institute's (NCI) SEER Program, which started collecting data from 1973, currently collects and publishes cancer incidence and survival data from population-based cancer registries covering approximately 26% of the U.S. population *(13,14)*. The National Program of Cancer Registries (NPCR), which was established by an act of

Congress in 1992, and which is administered by the CDC, supports central cancer registries in 45 states, the District of Columbia, and other U.S. territories *(15)*. These data represent 96% of the U.S. population. Together, NPCR and SEER collect data for the entire U.S. population.

3.1. SEER Database

The SEER Program is one of the best sources of information for evaluating cancer epidemiology in the United States. SEER currently collects and publishes data on patient demographics, primary tumor site, tumor morphology and stage at diagnosis, first course of treatment, and follow-up for vital status from 18 geographical areas in the United States, and it is considered representative of the U.S. population including minorities. SEER Program registries cumulatively have information on 6 million cancer cases, and >350,000 cases are added to the database each year. This database is updated annually and provided free of charge as a public service in print and web-based electronic formats, for use by researchers, physicians, public health officials, policy makers, and the public. The SEER program is considered to be the gold standard for all population based cancer registries with rigorous quality control measures.

3.2. SEER-Medicare Database

To provide patient-level information on different types of cancer in the United States, in a collaborative effort of the NCI, SEER registries, and the Centers for Medicare and Medicaid Services (CMS), the SEER database have been linked to claims-based measures of co-morbidities; screening and evaluation tests; and detailed treatment and treatment outcomes data, including cost data from the CMS *(7,16)*. Using a matching algorithm based on unique patient identifiers, such as social security numbers and date of birth, cancer data on individual patients available from the SEER registries was linked to a master Medicare enrollment first in 1991; and since then, it has been updated in 1995, 1999, and 2003. The SEER-Medicare data are available to outside investigators for research purposes. The SEER data included as part of the SEER-Medicare files are in a customized file known as the Patient Entitlement and Diagnosis Summary File. This file contains one record per person for individuals in the SEER data who have been matched with Medicare enrollment records with clinical information available for up to 10 diagnosed cancer cases, selected variables pertaining to Medicare enrollment information for that patient, and information about the median household economic and educational status for the census tract or zip code where the person resides. In general, all Medicare files have fields for race, sex, and date of birth or age, the date(s) of service, diagnostic codes, and procedure codes (either International Classification of Diseases [ICD]-9 codes for procedures and diagnoses or Health Care Financing Administration Common Procedure

Coding System codes for procedures) in addition to the amounts for charges and reimbursement. In addition, every Medicare file contains a provider identification number for the hospital or physician. Medicare files included as part of the SEER-Medicare data contain the SEER case number on each claim, which is the unique nonidentifiable number assigned to each cancer patient by the registries. To allow comparison studies with a control group without cancer, there are Medicare files containing similar information for a random sample of 5% of Medicare beneficiaries residing in the SEER areas persons who do not have cancer. The availability of the SEER-Medicare linked data provides researchers with a unique resource for extracting information on cancer with a patient-level focus and a longitudinal perspective before, during and after diagnosis of a particular cancer.

4. Overview of Statistical Techniques of Analyzing Cancer Registry Databases

Statistical analysis of cancer registry databases can be broadly grouped into two categories: descriptive and analytical. Traditionally, one of the basic functions of cancer registries have been to provide the public, scientists, researchers, and policy makers with descriptive data on incidence, mortality, prevalence, and survival of different cancers.

4.1. Incidence

Cancer incidence rate, a basic index of cancer epidemiology, is the number of newly diagnosed cancers of a specific site/type occurring in a specified population during a defined period. Age-adjusted incidence rates, and also temporal trends in incidence, are commonly derived statistics from population registry-based data. An age-adjusted rate is the weighted average of the age-specific rates where the weights are the proportion of persons in the corresponding age-specific group of a standard population. It should be noted that the unique characteristics of registry-based data require specialized statistical techniques for correct analysis and interpretation. For example, there is always an inevitable delay of certain period between the diagnosis of the cancer and its eventual reporting to a cancer registry. To adjust the current case count with the anticipated future corrections considering delay in reporting, the delay distribution of cancer cases has been modeled in precisely determining the current cancer trends, and also in monitoring the timeliness of cancer data collection by the registries. It has been shown that ignoring reporting delay and reporting error may produce downwardly biased cancer incidence trends, particularly in the most recent years of diagnosis *(17,18)*.

A risk-adjusted incidence rate can be calculated from registry databases that use first primary cancer as the numerator and the population who never had that particular cancer as denominator and help in understanding the actual transition rate of a healthy population to the cohort with a particular cancer *(19)*. Not surprisingly these risk adjusted rates are often different than the standard incidence rates that are derived from reported count of multiple instances of primaries of the same cancer in the numerator, and use the total population as the denominator.

Calculation of population-based measures of lifetime and age-conditioned probability of developing cancer, and also dying of cancer in the general population from a particular cancer, has been extensively described and is being increasingly used as practical and easily interpretable measures of cancer epidemiology *(20,21)*. The identification of changes in the temporal trend is an important issue in the analysis of cancer mortality and incidence data. Given the limitations of traditional linear and Poisson regression models, newer statistical techniques such as joinpoint models have been described to analyze registry-based data for temporal trends. The point in time where a trend changes direction is called a joinpoint. The joinpoint regression model describes continuous changes in rates, and by using a grid-search method, it fits a series of joined straight lines on a log scale to the expected annual percentage change in the incidence rate of a particular cancer over a defined number of years to fit the regression function with a number of joinpoints. Commonly, it uses a Monte Carlo permutation-based significance testing to determine the points in time when the direction of trends changes significantly *(22)*.

Patients with a first primary cancer are more likely than the average person to develop a subsequent malignancy, because of genetic susceptibility, a shared etiology, or even as a consequence of treatment of the first cancer. Also, effective screening and treatment regimens, coupled with more cancer diagnoses because of an aging population in the western world, have resulted in increasing numbers of cancer survivors who are at risk for subsequent cancers *(23)*. Cancer registry databases provide a unique opportunity to study the association of multiple primary cancers and to test hypotheses that explore plausible links in the etiology of different cancers, such as effect of smoking. The statistical methods used to investigate multiple primary malignant neoplasms in large populations, such as the SEER cohort, are well established *(24,25)*. A defined cohort of persons previously diagnosed with a certain cancer is followed through time to compare their subsequent cancer experience to the number of cancers that would be expected based on incidence rates for the general population. The standardized incidence ratio (SIR) is calculated as the ratio of the observed number of second primary malignancies to the expected number of second primary malignancies. The

statistical significance is usually assessed based on the assumption that the observed number of cases follows a Poisson distribution. In examining any two primary malignancies (A and B), two relevant statistical parameters must be evaluated: the SIR of A following B (SIR A/B) and the SIR of B following A (SIR B/A). Biologic plausibility of a significant association between a pair of primary malignancies is better established if the association is bidirectional. Mathematical modeling has demonstrated that under relatively general assumptions regarding the number of common risk factors, the prevalence of these factors, and the interaction (synergism) between them, the two SIRs should be nearly equal, provided that the lifetime risk profiles of the individuals in the study do not change.

4.2. Mortality

Mortality rates are another group of basic statistical indices commonly reported by analysis of cancer registry-based data. A cancer mortality rate is the number of reported cancer deaths of a specific site or type occurring in a specified population during a year (or group of years), usually expressed as the number of cancers per 100,000 population at risk. Several statistical methods and software tools have been developed for the analysis and reporting on cancer mortality statistics from cancer registry databases. These methods include age-adjusted rates with gamma confidence intervals *(26)*, trends in rates over time based on frequencies (such as percentage of change, annual percentage of change), and incidence-based mortality, which allows a description of mortality by selected variables associated with the cancer onset *(27)*.

4.3. Prevalence

Prevalence of cancer represents new and preexisting cases alive on a certain date in contrast to incidence, which reflects new cases of cancer diagnosed during a given period. Thus, prevalence is a function of both the incidence and survival. In cancer epidemiology, prevalence of a cancer is a statistical parameter of utmost importance because it truly reflects the burden of a particular cancer in the population, and it is important for policy makers in making decisions regarding healthcare resource allocation. One of the emerging prevalence measures is care prevalence, which is the measure of prevalent cases under care *(28)*. Although cancer registries typically do not provide such information, with the availability of linked databases (such as, the SEER-Medicare–linked database) allowing longitudinal tracking of patients with cancer, such measures are being increasingly reported as a more refined quantification of the burden of cancer. Another nontraditional measure of cancer prevalence is noncure prevalence, which is an estimate of prevalent cases that have not been cured of disease, and it may be a more specific and practical prevalence measure in terms of economics of cancer care *(29,30)*.

Prevalence of a particular cancer may be described as limited duration or complete prevalence *(31,32)*. Limited duration prevalence represents the proportion of people alive on a certain day that had a diagnosis of the disease within a defined period of past years. Complete prevalence represents the proportion of people alive on a certain day that previously had a diagnosis of the disease, regardless of how long ago the diagnosis was, or whether the patient is still under treatment or is considered cured. Registries of shorter duration (such as the SEER) with <40 or 50 years of data collection, can only estimate limited duration prevalence. In the United States, the only registry with sufficient length of follow-up data (since at least 1940) for reasonable prediction of complete cancer prevalence is the Connecticut Tumor Registry *(33)*. However, it should be noted that projecting estimates of national prevalence from a single regional registry has inherent issues with representativeness.

To derive complete prevalence from data from registries of shorter duration, a statistical modeling technique known as completeness index has been described and used by cancer registries, including SEER, in reporting the cancer statistical reviews *(30,34)*. Completeness index, which has been validated by SEER for selected cancer sites by using the Connecticut cancer registry, uses an estimation technique where complete prevalence is described as function of the observation time of a particular registry; incidence and survival indices before 1975 are predicted by modeling the SEER data. Advantages of the modeled completeness index are its stability even for rare cancers and importantly, that it permits estimation by SEER-derived race and ethnicity data. One of its disadvantages is that this technique generally cannot be applied for estimating complete prevalence of childhood cancers. Other approaches to estimate complete prevalence are cross-sectional population surveys *(35)*, the transition method rate *(36)*, and back calculation *(37,38)*. Cross-sectional population surveys use self-reporting for identification of cancer cases, but they obviously have limitations of underreporting and misclassification of cancer. A second approach is to use data from disease registries to estimate the various intensity (hazard or transition rate) functions that determine point prevalence *(36)*. As described by Keiding *(39)*, a person at calendar time t in the healthy state H may transit to the chronic disease state (e.g., cancer), I, with intensity a(t, z) that may depend on calendar time t or age z. Alternatively, the individual may die (state D) with intensity p(t, z) directly from state H. A person in state I is at risk of death with intensity X (t, x, d), which may depend on duration d in state I as well as on t and x. These intensities determine the prevalence of the chronic disease if one assumes hat the numbers of births at calendar time t is governed by process with intensity p(t) that is independent of the subsequent life histories. Derived from earlier statistical modeling

techniques applied for estimation of incidence and prevalence of chronic diseases, such as human immunodeficiency virus, techniques using back calculation methods have been described that allow estimation and projection of cancer prevalence patterns by using cancer registry incidence and survival data. As a first step, the method involves the fit of incidence data by an age, period, and cohort model to derive incidence projections. Prevalence is then estimated from modeled incidence and survival estimates. Cancer mortality is derived as a third step from modeled incidence, prevalence, and survival *(37, 38)*.

The counting method, which is commonly used to estimate prevalence, uses tumor registry data to count cases alive on a particular prevalence date, whereas adjustments are made to account for cases that are lost to follow-up who would otherwise have made it to the prevalence date. The expected number of cases lost to follow-up who make it to the prevalence date is computed using conditional survival curves for specified cohorts. Depending on the research question in the counting method, it is important to clarify which method was adopted in counting the tumor; often, only the first malignant primary tumor recorded in a particular registry is counted; in other instances, the first malignant tumor per site in a defined observation period is counted *(32,36)*.

4.4. Survival

Cancer survival is the proportion of patients alive at some point subsequent to the diagnosis of their cancer, or from some point after diagnosis (conditional survival). It is usually represented as the probability of a group of patients "surviving" a specified amount of time. It is important to understand that unlike incidence or mortality parameters, where the total population constitutes the denominator, only patients diagnosed with cancer are taken into account in calculating the survival parameter. Commonly used survival measures are observed all cause survival, and net cancer-specific survival, which is the probability of surviving cancer in the absence of other causes of death. Net cancer-specific survival rate does not take into account the impact of mortality from other causes; thus, it is considered a better parameter to understand temporal trends in survival or comparing survival in different racial and ethnic groups or even amongst different registries *(40)*. Conversely, crude probability of death (which is the probability of dying of cancer in the presence of other causes of death) is a better measure to assess cancer survival at a patient-level focus because mortality from other causes practically does play an important role in determining cancer survival for an individual patient.

Net cancer-specific survival and crude probability of death have two methods in which they can be estimated: using cause of death information or expected survival tables. Cause of death information in the registry data typically comes from death

certificates, which are often incorrect *(41)*. For example, in a patient dying from a metastatic cancer, the death certificate often cites the metastatic cancer is the cause of death rather than the primary cancer. One way of circumventing the errors inherent in the death certificates is to use expected survival rates from population based life expectancy tables, with an assumption that the general population dies of causes other than cancer at the same rate as the cancer population *(42,43)*. Besides being a relatively strong assumption, this method may be problematic in that such tables may not be available for defined cohorts in a particular geographic area. If life tables are used for estimating survival measures, then there are two basic measures. One measure, relative survival, is defined as the ratio of the proportion of observed survivors (all causes of death) in a cohort of cancer patients to the proportion of expected survivors in a comparable cohort of cancer-free individuals. The formulation is based on the assumption of independent competing causes of death. Because a cohort of cancer-free individuals is difficult to obtain, practically expected life tables are used assuming that the cancer deaths are a negligible proportion of all deaths. The second measure is crude probability of death by using expected survival, which uses expected survival (obtained from the expected life tables) to estimate the probability of dying from other causes in each interval.

There are a few issues that are important to understand in using survival estimates based on cancer registry data. There is always a lag between current year and available follow-up data in a particular registry, and it is important to specify the cohort of patients in terms of year of diagnosis and length of available follow-up data when making survival estimates. The other inherent issue in estimating long-term survival is that such estimates are only available for only those cohorts who were diagnosed a long time ago and have enough follow-up. Thus, direct estimates of long-term survival may not be very relevant for newly diagnosed patients, especially with respect to those cancers where there have been tremendous improvements in management and survival *(44)*. To provide more up-to-date estimates of survival for newly diagnosed patients, in the projection method that has been developed by SEER, a regression model is fit to interval relative survival and includes a parameter associated with a trend on diagnosis year *(45,46)*. The cumulative relative survival rate in a target year is calculated by multiplying the projected interval survival rates for that year.

4.5. Spatial

One of the newest applications of population-based cancer registry data is spatial representation and analysis of cancer data by using geographic information system (GIS). *(47,48)*. Such spatial representation or mapping of cancer data is becoming an invaluable tool in the exploration, analysis, and communication

of cancer data in understanding relationships between cancer and other health, socioeconomic, and environmental variables; and importantly, in enhancing computer- and Internet-based public health education *(49)*.

5. Statistical Software for Analyzing Cancer Registry Databases

The vast amount of data available in cancer registries and the complexities of the statistical analytical techniques involved require use of dedicated statistical software for any meaningful extraction and analysis of registry data. Fortunately, the SEER Program has taken a lead role in developing these statistical software packages, and several very useful software packages are now available. The foremost of the analytical software is the powerful SEER*Stat, which enables researchers analyze the entire SEER database and compute data on frequency distribution, incidence rates, temporal trends, and survival rates, including conditional survival *(50)*. Also, advanced statistical measures, such as limited duration prevalence, incidence-based mortality, and standardized incidence ratio for multiple primary cancers, can be calculated using this software. There is an accompanying software called The SEER*Prep software, which converts ASCII text data files to the SEER*Stat database format, thus allowing researchers analyze data from other cancer registries by using the SEER*Stat program *(51)*. The DevCan software computes probabilities of developing or dying from cancer for a hypothetical population for specific cancers, by using robust statistical techniques; to derive these probabilities, population estimates of incidence rates are obtained using cross-sectional counts of incident cases from the standard areas of the SEER Program and mortality counts for the same areas from data collected by the National Center for Health Statistics *(52)*. Joinpoint is a statistical software for the analysis of trends by using joinpoint models, and it takes trend data (e.g., cancer rates) and fits the simplest joinpoint model that the data allow *(53)*. ComPrev software estimates incidence and survival models using SEER cancer data for specific cancer sites, sex, and races to calculate the completeness index *(54)*. Complete prevalence is calculated by dividing limited duration prevalence by the completeness index as a proportion. Limited duration prevalence statistics can be generated from SEER*Stat software and imported into ComPrev. ProjPrev is a software made available by the SEER Program that is primarily used to derive U.S. prevalence by projecting SEER prevalence onto U.S. populations *(55)*. CanSurv is a powerful statistical software for analyzing population-based survival data *(56)*. For grouped survival data, it can fit both the

standard survival models and the mixture cure survival models, and it provides various graphs for model diagnosis. It also can fit parametric (cure) survival models to individually listed data. For geospatial analysis, Headbang is a software based on a smoothing algorithm for identifying a geographical patter by using the SEER data *(57)*. SaTScan is another software that allows to test for randomness of space distribution, time distribution, or both of a particular cancer, and it allows recognition of disease (cancer) clusters *(58)*. Most of these software packages are freely available for downloading from the SEER website, with appropriate tutorials, and they are backed by technical support by the SEER team. Besides SEER, other internet interfaces, such as CDC (state Cancer profiles) *(59)*, CINA + online Cancer in North America by NAACCR *(60)*, and Globocan by the IARC *(61)*, are frequently used by researchers to access aggregated cancer surveillance data generated from population-based cancer registries.

6. Impact of Cancer Registry Databases on Cancer Epidemiology

Cancer surveillance research based on cancer registry databases have over the past few decades expanded from its primary purview of descriptive epidemiology and disease quantification to a much broader analytical range, and it has made a significant contribution to every aspect of cancer epidemiology *(62)*. One direct and often underappreciated impact of cancer registry databases on cancer epidemiology is development of sophisticated and robust statistical techniques for solution in quantitative problems in cancer surveillance and control, population risk assessment, and development of methodology and relevant software for analyzing large databases with relative ease. The most visible and widely disseminated contributions of cancer registry databases remain descriptive periodic publications, such as the SEER Cancer Statistics Review, published annually by the Cancer Statistics Branch of the NCI; the annual report to the Nation on the Status of Cancer 1975–2003, jointly developed by several agencies; and internationally, the monograph on Cancer Incidence in Five Continents, published every fifth year by the IARC. These publications are statistical summaries that track the trend in cancer incidence, prevalence, survival, and mortality, and they serve as the most recognized references on cancer epidemiology globally. More detailed analysis of cancer registries have identified patterns that often point to specific etiologies. Landmark examples include the identification of perimenopausal shift in the age-specific incidence curve of breast cancer in women, implicating reproductive and hormonal factors in

etiology *(63)*; different rates of gastric cancer in second generation and immigrant U.S. Japanese, highlighting the interaction of genetic and environmental factors *(64)*; the role of pesticides in prostrate cancer as revealed in the Agricultural Health Study *(65)*; excess bladder cancer risk in truck drivers, workers exposed to motor exhaust, and workers within the chemical, rubber, and plastics industries, as suggested by the National Bladder Cancer Study *(66)*; several studies pointing to the carcinogenic effects of environmental tobacco smoke *(67)*; the relationship between oral contraceptive and menopausal estrogen use and breast, endometrial, and ovarian cancers among U.S. women, as studied in the Cancer and Steroid Hormone Study *(68)*; and role of diet *(69)*, physical activity *(70)*, and nonsteroidal anti-inflammatory drug use *(71)* in many common cancers, such as colon and reproductive cancers. A special example in studying causality of cancer is study of multiple primary cancers, which is practically impossible without using large cancer registry databases, and many studies have used these databases to study the association of multiple primary cancers and plausible etiological factors, such as shared risk or exposure, effect of treatment, or genetic susceptibility *(72–76)*. Familial cancer registries are rapidly emerging to be a powerful tool in studying genetic susceptibility and in identifying genetic factor in the etiopathogenesis of many common cancers, such as esophageal, pancreatic, colon, and breast cancers. The infrastructure of cancer registry databases such as the SEER has been critical both to the recruitment of these families and to the retrieval of related cancer data for conducting large multicenter, population-based studies such as the Women's Environment, Cancer, and Radiation Epidemiology study that is investigating gene–environment interactions that may influence susceptibility to breast cancer *(77–79)*. Surveillance data that form the foundation of the cancer registry databases provide unique glimpse into different aspects of cancer epidemiology. Important examples of such high-impact studies that were primarily derived from careful analysis of cancer registry databases included recent recognition of rising trend in the incidence of esophageal and gastroesophageal junctional adenocarcinoma in the western countries *(80)*; the association of Kaposi's sarcoma in patients with AIDS *(81)*, use of the GIS techniques in identifying statistically significant increase in childhood cancers for the period 1979 to 1995 in Dover Township in New Jersey, possibly related to exposure to environmental carcinogen *(82)*; and identification of geographically clustered neighborhoods with high rates of late-stage breast cancers *(83)*. Cancer registry databases also are unique resources for studying epidemiology of rare cancers such as the male breast cancer *(84)* and also cancer in small population groups such as the native Americans *(85)*.

Cancer registry-based analytical studies are becoming increasingly important in public health education, and in initiating and evaluating public health efforts. The lifetime risk of developing breast cancer, which is a commonly cited statistic and has been extensively used for health education and advocating targeted health interventions, was derived from studies conducted using the SEER database *(86,87)*. Health disparities in the minorities and socioeconomically deprived sections of the population is a stark, unpleasant reality, and researchers increasingly use the cancer registry databases to identify and highlight such disparities to bring about appropriate public heath and sociopolitical interventions *(88)*. Important studies based on the cancer registry database have highlighted the influence of socioeconomic factors on cancer incidence, treatment outcomes, and mortality by using the linkage of cancer registry data to other databases, such as the U.S. census, which provided information on selected socioeconomic variables at the neighborhood level *(89–93)*. Data from cancer registry databases also provide the final yardstick for measuring the long-term impact of implementation of public policy, as was demonstrated by the study that demonstrated rapid decline of lung cancer in association with the California Tobacco Control program *(94)*.

The SEER-initiated "pattern of care" studies effectively demonstrate the potential of cancer registry databases in studying complex issues in the area of treatment outcomes and end results, which are increasingly becoming an integral concept in understanding the epidemiology of cancer in its broadest purview *(95–97)*. Similarly, linkage with the Medicare database offers researcher innovative use of cancer registry databases in studying issues as diverse as cancer control practices and their effect on the cancer burden; patterns of access to cancer care; impact of co-morbidities, race, geographic, socioeconomic, and provider-related factors on access, diagnosis, treatment, and treatment outcomes; and importantly, cost-effectiveness of cancer care *(7,98–101)*. Despite the limitations of training and funding opportunities *(62)*, since 1974 >4,500 scientific publications have been published using the SEER and other linked databases in the U.S. alone, leaving no doubt regarding the enormity of the impact of cancer registry databases in cancer epidemiology.

In conclusion, over the past few decades cancer registry databases have evolved a long way in terms of number, coverage, technical sophistication, quantity, quality, and scope of information, and they are increasingly recognized as an indispensable part of local, national, and international cancer control programs. It is certain that cancer registry databases will continue to be an extraordinary resource of information for researchers, scientists, policy makers, and the public in our uphill and global fight against cancer.

References

1. Clive, R. E. (2004) Introduction to cancer registries, in Cancer Registry Management: Principles and Practice (Hutchison, C. L., ed.), Kendall Hunt Publishing, Dubuque, IA, pp. 1–9.
2. Parkin, D. M. (2006) The evolution of the population-based cancer registry. Nat. Rev. Cancer 6(8), 603–12.
3. Parkin, D. M., Chen, V. W., Ferlay, J., Galceran, J., Storm, H. H., and Whelan, S. L. (1994) Comparability and quality control in cancer registration, in IARC Technical Report No. 19, International Agency for Research on Cancer, Lyon, France.
4. Jensen O. M. (ed.) (1991) Cancer Registration, Principles and Methods, No. 95, International Agency for Research on Cancer, Lyon, France.
5. Working Group of the International Association of Cancer Registries (2005) Guidelines for confidentiality in population-based cancer registration. Eur J Cancer Prev **14,** 309–327.
6. National Cancer Institute (2007) SEER statistical resources. http://seer.cancer.gov/resources/. Cited 4 April 2007.
7. Warren, J. L., Klabunde, C. N., Schrag, D., Bach, P. B., and Riley, G.F. (2002) Overview of the SEER-Medicare data, content, research applications, and generalizability to the United States elderly population. Med. Care **40,** 5–18.
8. Spittle, M. F. (1998) Spectrum of AIDS-associated malignant disorders. Lancet **351(9119),** 1833–9.
9. Centers for Disease Control and Prevention (2007) National Center for Chronic Disease Prevention and Health Promotion. Behavioral Risk Factor Surveillance System. http://www.cdc.gov/brfss/index.htm. Cited 4 April 2007.
10. National Center for Health Statistics (2007) National Health Interview Survey. http://www.cdc.gov/nchs/nhis.htm. Cited 4 April 4 2007.
11. National Center for Health Statistics (2007) National Hospital Discharge and Ambulatory Surgery Data. http://www.cdc.gov/nchs/about/major/hdasd/nhds.htm. Cited 4 April 2007.
12. National Center for Health Statistics (2007) Ambulatory Health Care Data. http://www.cdc.gov/nchs/about/major/ahcd/ahcd1.htm. Cited 4 April 2007.
13. Harlan, L.C., and Hankey, B.F. (2003) The surveillance, epidemiology, and end-results program database as a resource for conducting descriptive epidemiologic and clinical studies. J Clin Oncol **21(12),** 2232–3.
14. Hankey, B. F., Ries, L. A., and Edwards, B. K. (1999) The surveillance, epidemiology, and end results program: a national resource. Cancer Epidemiol Biomarkers Prev **8(12),** 1117–21.
15. Centers for Disease Control and Prevention (2007) National Program of Cancer Registries. http://www.cdc.gov/cancer/npcr/about.htm. Cited 4 April 2007.
16. National Cancer Institute (2007) SEER-Medicare Linked Database. http://health-services.cancer.gov/. Cited 4 April 2007.
17. Clegg, L. X., Feuer, E. J., Midthune, D., Fay, M. P., and Hankey, B. F. (2002) Impact of reporting delay and reporting error on cancer incidence rates and trends. J Natl Cancer Inst **94,** 1537–45.
18. Midthune, D. N., Fay, M. P., Clegg, L. X., and Feuer, E. J. (2005) Modeling reporting delays and reporting corrections in cancer registry data. J Am Stat Assoc **100(469),** 61–70.
19. Merrill, R. M., and Feuer, E. J. (1996) Risk-adjusted cancer-incidence rates (United States). Cancer Causes Control **7(5),** 544–52.
20. Fay, M. P. (2004) Estimating age conditional probability of developing disease from surveillance data. Popul Health Metr **2(1),** 6.
21. Fay, M. P., Pfeiffer, R., Cronin, K. A.., Le, C., and Feuer, E. J. (2003) Age-conditional probabilities of developing cancer. Stat Med **22(11),** 1837–48.
22. Kim, H. J., Fay, M. P., Feuer, E. J., and Midthune, D. N. (2000) Permutation tests for joinpoint regression with applications to cancer rates. Stat Med **19,** 335–51
23. Curtis, R. E., Freedman, D. M., Ron, E., Ries, L. A. G., Hacker, D. G., Edwards, B. K., Tucker, M. A., and Fraumeni, J. F., Jr. (eds) (2006) New malignancies among cancer survivors: SEER Cancer Registries, 1973–2000. National Cancer Institute. National Institutes of Health Publ. No. 05-5302. National Institutes of Health, Bethesda, MD.
24. Schoenberg, B. S., and Myers, M. H. (1977) Statistical methods for studying multiple primary malignant neoplasms. Cancer **40,** 1892–1898.
25. Begg, C. B., Zhang, Z., Sun, M., Herr, H. W., and Schantz, S. P. (1995) Methodology for

evaluating the incidence of second primary cancers with application to smoking-related cancer from the Surveillance, Epidemiology, and End Results (SEER) Program. Am J Epidemiol **142,** 653–665.

26. Fay, M. P., and Feuer, E. J. (1997) Confidence intervals for directly standardized rates: a method based on the Gamma distribution. Stat Med **16,** 791–801.
27. Chu, K. C., Miller, B. A., Feuer, E. J., and Hankey, B. F. (1994) A method for partitioning cancer mortality trends by factors associated with diagnosis: an application to female breast cancer. J Clin Epidemiol **47(12),** 1451–61.
28. Mariotto, A., Warren, J. L., Knopf, K. B., and Feuer, E. J. (2003) The prevalence of colorectal cancer patients under care in the US. Cancer **98,** 1253–61.
29. Coldman, A. J., McBride, M. L., and Braun, T. (1992) Calculating the prevalence of cancer. Stat Med **11,** 1579–89.
30. Capocaccia, R., and De Angelis, R. (1997) Estimating the completeness of prevalence based on cancer registry data. Stat Med **16,** 425–40.
31. Clegg, L., Gail, M., and Feuer, E. J. (2002) Estimating the variance of disease prevalence estimates from population-based registries. Biometrics **58(3)**, 684–8.
32. Feldman, A. R., Kessler, L., Myers, M. H., and Naughton, M. D. (1986) The prevalence of cancer. N Engl J Med **315,** 1394–7.
33. Gershman, S. T., Flannery, J. T., Barrett, H., Nadel, R. K., and Meigs, J. W. (1976) Development of the Connecticut Tumor Registry. Conn Med **40,** 697–701.
34. Merrill, R. M., Feuer, E. J., Cappacaccia, R., and Mariotto, A. (2000) Cancer prevalence estimates based on tumor registry data in the SEER Program. Int J Epidemiol **29,** 197–207.
35. Byrne, J., Kessler, L. G., and Devesa, S. S. (1992) The prevalence of cancer among adults in the United States. Cancer **68,** 2154–9.
36. Gail, M. H., Kessler L, Midthune, D., and Scoppa S. (1999) Two approaches for estimating disease prevalence from population-based registries of incidence and total mortality. Biometrics **55,** 1137–44.
37. De Angelis, G. D., De Angelis, R., Frova, L., and Verdecchia, A. (1994) MIAMOD: a computer program to estimate chronic disease morbidity using mortality and survival data. Comput Methods Programs Biomed **44,** 99–107.
38. Verdecchia, A., De Angelis, G., and Capocaccia, R. (2002) Estimation and projections of cancer prevalence from cancer registry data. Stat Med **21(22),** 3511–26.
39. Keiding, N. (1991) Age-specific incidence and prevalence: a statistical perspective. J R Stat Soc Ser A **154,** 371–412.
40. Marubini, E., and Valsecchi, M. G (eds) (2004) Analyzing Survival Data from Clinical Trials and Observational Studies, Wiley, Hoboken, NJ.
41. Boer, R., Ries, L., van Ballegooijen, M., Feuer, E., Legler, J., and Habbema, D. (2003) Ambiguities in calculating cancer patient survival: the SEER experience for colorectal and prostate cancer. Technical Report 2003-05, Statistical Research and Applications Branch, National Cancer Institute, Bethesda, MD (http://srab.cancer.gov/reports).
42. Brown, C.C. (1983) The statistical comparison of relative survival rates. Biometrics **39,** 941–8.
43. Ederer, F., Axtell, L. M., and Cutler, S. J. (1961) The relative survival rate: a statistical methodology. Natl Cancer Inst Monogr **6,** 101–21.
44. Brenner, H., and Hakulinen, T. (2002) Advanced detection of time trends in long-term cancer patient survival: experience from 50 years of cancer registration in Finland. Am J Epidemiol **156(6),** 566–77.
45. Yu, B., Tiwari, R. C., Cronin, K. A., McDonald, C., and Feuer, E. J. (2005) CANSURV: a Windows program for population-based cancer survival analysis. Comput Methods Programs Biomed **80(3),** 195–203.
46. Yu, B., Tiwari, R. C., Cronin, K. A, and Feuer, E. J. (2004) Cure fraction estimation from the mixture cure models for grouped survival data. Stat Med **23(11),** 1733–47.
47. Brewer, C. A. (2006) Basic mapping principles for visualizing cancer data using geographic information systems (GIS). Am J Prev Med **30(2 Suppl),** S25–S36.
48. Bell, B. S, Hoskins, R. E., Pickle, L. W., and Wartenberg, D. (2006) Current practices in spatial analysis of cancer data: mapping health statistics to inform policymakers and the public. Int J Health Geogr **8,** 5:49.
49. Pickle, L. W., Hao, Y., Jemal, A., Zou, Z, Tiwari, R. C., and Ward, E., et al. (2007) A new method of estimating United States and state-level cancer incidence counts for the current calendar year. CA Cancer J Clin 57, 30–42.
50. National Cancer Institute (2006) SEER*Stat software, version 6.2. Statistical Research

and Applications Branch, National Cancer Institute. http://seer.cancer.gov/seerstat/. Cited 4 April 2007.

51. National Cancer Institute (2006) SEER*Prep software, version 2.3.5. Statistical Research and Applications Branch, National Cancer Institute. http://srab.cancer.gov/seerprep/. Cited 4 April 2007.
52. National Cancer Institute (2006) DevCan 6.1.1 for Windows software, version 6.1.1. Statistical Research and Applications Branch, National Cancer Institute. http://srab.cancer.gov/devcan/. Cited 4 April 2007.
53. National Cancer Institute (2006) Joinpoint regression program, version 3.0. Statistical Research and Applications Branch, National Cancer Institute. http://srab.cancer.gov/joinpoint/. Cited 4 April 2007.
54. National Cancer Institute (2006) Complete Prevalence (ComPrev) software, version 1.0. Statistical Research and Applications Branch, National Cancer Institute. http://srab.cancer.gov/comprev/. Cited 4 April 2007.
55. National Cancer Institute (2006) Projected Prevalence (ProjPrev) software, version 1.0.1. Statistical Research and Applications Branch, National Cancer Institute. http://srab.cancer.gov/projprev/. Cited 4 April 2007.
56. National Cancer Institute (2006) CanSurv software, version 1. Statistical Research and Applications Branch, National Cancer Institute. http://srab.cancer.gov/cansurv/. Cited 4 April 2007.
57. National Cancer Institute (2006) HeadBang PC software, version 3.0. Hansen; Simonson and Statistical Research and Applications Branch, National Cancer Institute. http://srab.cancer.gov/headbang/. Cited 4 April 2007.
58. Kulldorff, M. and Information Management Services, Inc. (2007) SaTScan™ version 7.0: software for the spatial and space-time scan statistics. http://www.satscan.org/. Cited 4 April 2007.
59. National Cancer Institute (200X) State cancer profiles. http://statecancerprofiles.cancer.gov/index.html/. Cited 4 April 2007.
60. North American Association of Central Cancer Registries (200X) CINA + Online Cancer in North America. http://www.cancer-rates.info/naaccr/. Cited 4 April 2007.
61. Ferlay, J., Bray, F., Pisani, P., and Parkin, D. M. (2004) GLOBOCAN 2002: Cancer Incidence, Mortality and Prevalence Worldwide. IARC Cancer Base No. 5. Version 2.0. IARC Press, Lyon, France. (http://www-dep.iarc.fr/).
62. Glaser, S. L., Clarke, C. A., Gomez, S. L., O'Malley, C. D., Purdie, D. M., and West, D. W. (2005) Cancer surveillance research, a vital subdiscipline of cancer epidemiology. Cancer Causes Control **16,** 1009–1019.
63. Clemmesen, J. (1948) Carcinoma of the breast. Results from statistical research (symposium). Br J Radiol **21,** 583–590.
64. Buell, P., and Dunn, J. E., Jr. (1965) Cancer mortality among Japanese Issei and Nisei of California. Cancer **18,** 656–64.
65. Alavanja, M. C., Sandler, D. P., McMaster, S. B., Zahm, S. H., McDonnell, C. J., and Lynch, C. F., et al. (1996) The Agricultural Health Study. Environ Health Perspect **104 (4),** 362–369.
66. Silverman, D. T., Hartge, P., Morrison, A. S., and Devesa, S. S. (1992) Epidemiology of bladder cancer. Hematol Oncol Clin North Am **6(1),** 1–30.
67. National Cancer Institute Health (1999) Effects of Exposure to Environmental Tobacco Smoke, Smoking and Tobacco Control Monograph No. 10. U.S. Department of Health and Human Services, National Cancer Institute, National Institutes of Health, Bethesda, MD.
68. No authors listed (1986) Oral contraceptive use and risk of breast cancer. The Cancer and Steroid Hormone Study of the Centers for Disease Control and the National Institute of Child Health and Human Development. N Engl J Med **315,** 405–411.
69. Le Marchand, L., Yoshizawa, C. N., Kolonel, L. K., Hankin, J. H., and Goodman, M. T. (1989) Vegetable consumption and lung cancer risk: a population-based case-control study in Hawaii. J Natl Cancer Inst **81,** 1158–1164.
70. Irwin, M. L., Aiello, E.J., McTiernan, A., Bernstein, L., Gilliland, F. D., Baumgartner, R. N., et al. (2007) Physical activity, body mass index, and mammographic density in postmenopausal breast cancer survivors. J Clin Oncol **25(9),** 1061–6.
71. Cerhan, J. R., Anderson, K. E., Janney, C. A., Vachon, C. M., Witzig, T. E., and Habermann, T. M. (2003) Association of aspirin and other non-steroidal anti-inflammatory drug use with incidence of non-Hodgkin lymphoma. Int J Cancer **106(5),** 784–8.
72. Kleinerman, R. A., Boice, J. D., Jr., Storm, H. H., Sparen, P., Andersen, A., Pukkala, E.,

Lynch, C. F., et al. (1995) Second primary cancer after treatment for cervical cancer. An international cancer registries study. Cancer **76,** 442–452.

73. Travis, L. B., Curtis, R. E., Storm, H., Hall, P., Holowaty, E., Van Leeuwen, F. E., et al. (1997) Risk of second malignant neoplasms among long-term survivors of testicular cancer. J Natl Cancer Inst **89,** 1429–1439.
74. Travis, L. B., Gospodarowicz, M., Curtis, R. E., Clarke, E.A., Andersson, M., Glimelius, B., et al. (2002) Lung cancer following chemotherapy and radiotherapy for Hodgkin's disease. J Natl Cancer Inst **94,** 182–192.
75. Zablotska, L. B., Chak, A., Das, A., and Neugut, A.I. (2005) Increased risk of squamous cell esophageal cancer after adjuvant radiation therapy for primary breast cancer. Am J Epidemiol **161(4),** 330–7.
76. Das, A., Neugut, A. I., Cooper, G. S., and Chak, A. (2004) Association of ampullary and colorectal malignancies. Cancer **100(3),** 524–30.
77. Brewster, D. H, Fordyce, A., and Black, R. J. (2004) Scottish clinical geneticists. Impact of a cancer registry-based genealogy service to support clinical genetics services. Fam Cancer **3(2),** 139–41.
78. Ziogas, A., and Anton-Culver, H. (2003) Validation of family history data in cancer family registries. Am J Prev Med **24(2),** 190–198.
79. Bernstein, J. L., Langholz, B. M., Haile, R.W., Bernstein, L., Thomas, D. C., and Stovall, M., et al. (2004) Study design: evaluating gene-environment interactions in the etiology of breast cancer—the WECARE Study. Breast Cancer Res **6,** R199–R214
80. Blot, W. J., Devesa, S. S., Kneller, R. W., and Fraumeni, J. F., Jr. (1991) Rising incidence of adenocarcinoma of the esophagus and gastric cardia. JAMA **265,** 1287–1289.
81. Reynolds, P., Layefsky, M. E., Saunders, L. D, George, F. L., and Payne, S. F. (1990) Kaposi's sarcoma reporting in San Francisco: comparison of AIDS and cancer surveillance systems. J Acquir Immune Defic Syndr **3(Suppl 1),** S8–S13.
82. Blumenstock, J., Fagliano, J., and Bresnitz, E. (2000) The Dover Township childhood cancer investigation. N J Med **97,** 25–30.
83. Roche, L. M., Skinner, R., and Weinstein, R. B. (2002) Use of a geographic information system to identify and characterize areas with high proportions of distant stage breast cancer. J Public Health Manag Pract **8,** 26–32.
84. Giordano, S. H., Cohen, D. S., Buzdar, A. U., Perkins, G., and Hortobagyi, G. N. (2004) Breast carcinoma in men: a population-based study. Cancer **101(1),** 51–7.
85. Swan, J., and Edwards, B. K. (2003) Cancer rates among American Indians and Alaska Natives: is there a national perspective. Cancer **98(6),** 1262–72
86. Gail, M. H., Brinton, L. A., Byar, D. P, Corle, D. K., Green, S. B., Schairer, C., et al. (1989) Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst **81,** 1879–1886.
87. Feuer, E. J., Wun, L. M., Boring, C. C., Flanders, W. D., Timmel, M. J., and Tong, T. (1993) The lifetime risk of developing breast cancer. J Natl Cancer Inst **85,** 892–897.
88. Harper, S., and Lynch, J. (2005) Methods for measuring cancer disparities: using data relevant to healthy people 2010 cancer-related objectives. NCI Cancer Surveillance Monograph Series, No. 6. National Cancer Institute, Bethesda, MD.
89. Humble, C. G., Samet, J. M., Pathak, D. R., and Skipper, B. J. (1985) Cigarette smoking and lung cancer in 'Hispanic' whites and other whites in New Mexico. Am J Public Health **75(2),** 145–148.
90. Chen, V. W., Fontham, E. T. H., Craig, J. F., Groves, F. D., Culley, P., Rainey, J. M., et al. (1992) Cancer in South Louisiana. Part I: tobacco-related cancers. J La Med Soc **144,** 149–155.
91. Friedell, G. H., Tucker, T. C., McManmon, E., Moser, M., Hernandez, C., and Nadel, M. (1992) Incidence of dysplasia and carcinoma of the uterine cervix in an Appalachian population. J Natl Cancer Inst **84,** 1030–1032.
92. Chen, V. W., Wu, X. C., Andrews, P. A., Fontham, E. T., and Correa, P. (1994) Advanced stage at diagnosis: an explanation for higher than expected cancer death rates in Louisiana? J La Med Soc **146,** 137–145.
93. Mills, P. K., and Kwong, S. (2001) Cancer incidence in the united farm workers of America (UFW) 1987–1997. Am J Ind Med **40,** 596–603.
94. Barnoya, J., and Glantz, S. (2004) Association of the California tobacco control program with declines in lung cancer incidence. Cancer Causes Control **15(7),** 689–95.
95. Harlan, L. C., Abrams, J., Warren, J. L., Clegg, L., Stevens, J., and Ballard-Barbash, R. (200) Adjuvant therapy for breast cancer:

practice patterns of community physicians. J Clin Oncol **20,** 1809–1817.

96. Potosky, A. L., Harlan, L. C., Kaplan, R. S., Johnson, K. A., and Lynch, C. F. (2002) Age, sex, and racial differences in the use of standard adjuvant therapy for colorectal cancer. J Clin Oncol **20,** 1192–1202.
97. Harlan, L. C., Clegg, L. X., and Trimble, E. L. (2003) Trends in surgery and chemotherapy for women diagnosed with ovarian cancer in the U.S. J Clin Oncol **21(18),** 3488–94
98. Potosky, A. L., Riley, G. F., Lubitz, J. D., Mentnech, R. M., and Kessler, L. G. (1993) Potential for cancer related health services research using a linked Medicare-tumor registry database. Med Care **31(8),** 732–748.
99. Brown, M. L., Riley, G. F., Schussler, N., and Etzioni, R. (2002) Estimating health care costs related to cancer treatment from SEER-Medicare data. Med Care **40(8 Suppl IV),** 104-17.
100. Doebbeling, B. N., Wyant, D. K., McCoy, K. D., Riggs, S., Woolson, R. F., Wagner, D., et al. (1999) Linked insurance-tumor registry database for health services research. Med Care **37(11),** 1105–15.
101. McClish, D. K., Penberthy, L., Whittemore, M., Newschaffer, C., Woolard, D., Desch, C. E., et al. (1997) Ability of Medicare claims data and cancer registries to identify cancer cases and treatment. Am J Epidemiol **145(3),** 227–33.

Chapter 3

Breast Cancer in Asia

Cheng-Har Yip

Summary

Breast cancer is the commonest cancer in most countries in Asia. The incidence rates remain low, although increasing at a more rapid rate than in western countries, due to changes in the lifestyle and diet. There are many differences between breast cancer in Asia compared with western countries. The mean age at onset is younger than in the west, and unlike the west, the age-specific incidence decreases after the age of 50 years. Because there is no population-based breast cancer screening program in the majority of Asian countries, the majority of patients present with advanced disease. There is a higher proportion of hormone receptor-negative patients, and some evidence that the cancers in Asia are of a higher grade. Most of the Asian countries are low- and middle-income countries, where access to effective care is limited. Because of the late detection and inadequate access to care, survival of women with breast cancer in Asia is lower than in western countries. Improving breast health in most of the Asian countries remains a challenge that may be overcome with collaboration from multiple sectors, both public and private.

Key words: Breast cancer, epidemiology, survival, histopathology, screening, treatment.

1. Introduction and Background

Asia is the world's largest and most populous continent; it contains 60% of the world's population. Except for a few countries (Singapore, Taiwan, Hong Kong, Japan, South Korea, Israel, Saudi Arabia, and Macau) that are classified as high-income countries, the rest of Asia includes low- and middle-income countries.

The two most important sites of female cancers across the countries of Asia, in terms of incidence, are the breast and the cervix uteri (**Table 3.1**). In some countries, such as South Korea, China, and Japan, stomach cancer is also an important cancer, whereas in India, Myanmar, and Thailand, cervical cancer is the commonest cancer in women. The highest incidence rates

Mukesh Verma (ed.), *Methods in Molecular Biology, Cancer Epidemiology, vol. 471*
© 2009 Humana Press, a part of Springer Science + Business Media, Totowa, NJ
Book doi: 10.1007/978-1-59745-416-2

Table 3.1
Incidence of cancer in females in selected countries in Asia (per 100,000)

Cancer site	China	South Korea	Japan	India	Pakistan	Malaysia	Singapore	Philippines	Indonesia	Iran	Saudi Arabia	Myanmar	Thailand
Lung	1.9	12.8	12.3	1.9	2.8	10.9	18.3	13.5	6.8	2.2	2.8	12.7	10.4
Breast	18.7	20.4	32.7	19.1	50.1	30.8	48.7	46.6	26.1	17.1	24.7	20.2	16.6
Stomach	19.2	26.8	26.1	2.8	2.7	6.2	11.1	5.2	2.1	11.1	3.5	7.8	2.9
Cervix	6.8	17.9	8.0	30.7	6.5	15.7	13.2	20.9	15.7	4.4	4.6	24.6	19.8

Source: **Ref.** *1.*

(age-standardized rates; ASRs) for breast cancer in Asia are in Pakistan (50.1 per 100,000), Singapore (48.7 per 100,000), and the Philippines (46.6 per 100,000), whereas the lowest rates are in Laos (10.9 per 100,000) and Mongolia (6.6 per 100,000) These values are compared with the ASRs in the United States (101.1 per 100,000), Uniting Kingdom (87.2 per 100,000), and Australia (83.2 per 100,000 *(1)*. Hence, although breast cancer is an important cancer in Asian women, the incidence rates remain much lower than in western populations. However, the incidence in Asia is increasing at a more rapid rate than in the west; in Singapore, a 3.6% increase per annum was noted over a 25-year period *(2)*.

There are many possible explanations for this increasing trend: earlier age at menarche, later age at menopause, decreasing fertility, increasing age at first birth, increases in height and weight, and changes in diet *(3)*. Migrant studies have shown that environmental and lifestyle factors play an important role, because studies of Japanese and Chinese migrants to the United States show a progressive increase in risk in successive generations *(4,5)*.

Even within the same country, there are differences in the incidence of breast cancer among different ethnic groups. In Malaysia, which is a multiracial country, the National Cancer Registry in 2003 reported an incidence of 59.7 per 100,000 in Chinese, 55.8 per 100,000 in Indians, and 33.9 per 100,000 in Malays *(6)*. The ethnic difference even within the same country implies that there are other factors that may be of importance in the etiology apart from geographical location. Although the Chinese and Indians are originally immigrants from their own country in the late 1800s and early 1900s, the three races have retained their distinctive culture and dietary practices, which could explain their differing risk of breast cancer.

2. Age Incidence of Breast Cancer in Asia

Although Asia is a vast continent comprising many countries, from low income (e.g., Vietnam) to low middle income (e.g., Indonesia) to upper middle income (e.g., Malaysia) and high-income (e.g., South Korea) countries, there are many similarities in the age of onset of breast cancer. Breast cancer occurs at a younger age in Asia. The mean age is around 50 years, and the prevalent age group is 40–49 years old. More than 50% of the patients are under the age of 50 years, and >60% are premenopausal *(7,8)*. The younger mean age can be explained by the young population pyramid structure of developing countries, which is broad based, compared with the wide-based population structure in developed countries.

Breast cancer incidence in Asia has a distinctive age-specific curve. The age-specific incidence increases rapidly until the age of 50, and then it continues to increase at a slower rate, probably due to diminishing levels of circulating estrogens *(9)*. However in low-incidence countries, the age-specific incidence decreases after menopause This decrease is a consequence of increasing risks of occurrence in successive generations of women born after World War II (birth cohort effect) rather than a real decline in risk with age *(10,11)*. With time, the age incidence in Asia could be similar to that in western countries. This trend is seen in Japan, where the mean age at diagnosis has increased from 48.0 years in 1946–1959 to 53.9 years in 2000–2001, and the percentage of postmenopausal breast cancer patients has increased from 40.6% in 1946–1959 to 55.4% in 2000–2001 *(12)*.

3. Stage at Presentation

Breast cancer presents at later stages in Asia compared with western countries *(13,14)*. The delay in presentation is attributed mainly to the various barriers that exist in the Asian region. Such barriers can be structural (e.g., poor health facilities, distance to health care facility, inability to take time off work) or organizational (e.g., difficulty in navigating the complex health care system and interaction with medical staff). Psychological and sociocultural barriers include poor health motivation, denial of personal risk, fatalism, mistrust of cancer treatments, and the fear of becoming a burden to family members. In some traditional cultures, especially the Muslim culture, a woman's decision and actions are controlled by men, and men may be unaware of breast screening. In some cultures in Asia, there is also the strong influence of traditional medicine *(15,16)*.

In India, 50–55% present with stages 3 and 4 breast cancer (late stage) *(17)*. In Kuala Lumpur, Malaysia, in a study in two institutions, 30–40% present with stages 3 and 4 in a teaching hospital, with a median size of 4.2 cm, whereas 50–60% present with stages 3 and 4 in the city general hospital, with a median size of 5.4 cm *(13)*. Even within the same country, there are differences in the stage at presentation between the different ethnic groups, where the Malays in Malaysia present with larger tumors and later stages compared with the Chinese and Indians *(7)*. In Oman, a study of 152 breast cancer patients from 1996 to 2002 showed that 50.1% presented with stages 3 and 4, with a mean tumor size of 4.6 cm *(18)*. A study in a single institution in Iran *(19)* showed that 61% of women presented with stages 3 and 4.

In contrast, in the few developed countries in Asia (e.g., Japan, South Korea, and Singapore), women present with earlier stages and with smaller tumors. In Korea, the percentage of early stages 0 and 1 breast cancer is 45.2%, with a screen-detected rate of 17.8% *(20)*. In Japan, 90% of patients presented with stages 0–2 breast cancer in 2000–2001, and 13% were carcinoma *in situ (11)*. In Singapore, 21.5% of women presented with locally advanced or metastatic breast cancer in a tertiary institution from 2000 to 2003 *(21)*.

These data show that with time, as the developing countries in Asia become more developed, with better education and access to health care, women should be presenting with earlier stages of breast cancer.

4. Histopathology of Breast Cancer in Asia

There are distinct differences in the histopathology of breast cancer in Asian women compared with Caucasian women. A study in different ethnic groups in the United States *(22)* showed that Asians have a higher nonstatistically significant odds of ductal versus lobular carcinoma compared with Caucasians, whereas a study in Japan *(23)* showed a lower incidence of lobular carcinoma in Japanese women compared with Caucasian women; 81.2% of breast cancers in Iran *(24)* and 85.6% of breast cancers in Malaysia *(25)* are infiltrating ductal carcinoma.

Hormone receptor status among the non-Hispanic whites and blacks in the United States differs significantly. In a study of 13,239 cases of breast cancer, Gapstur et al. *(26)* showed that in non-Hispanic whites, 74% of breast cancers were estrogen receptor (ER) positive, 65% were progesterone receptor (PgR) positive, and 80% were ER or PgR positive, whereas in blacks, 58% were ER positive, 51% were PgR positive, and 65% were ER or PgR positive. The hormone receptor status is Asians also differs from that in Caucasians. A study in the United States showed that Asians had a 1.4- to 3.1-fold elevated risk of presenting with ER-negative and PgR-negative breast cancer *(27)*. In India, it was initially reported that only 32.6% of Indian women with breast cancer were ER positive, 46% were PgR positive, and there was an abnormally high percentage (21.1%) of ER negative, PgR positive rather than the usual 3–5% reported in Caucasians *(28)*. However, a later study from India, using automated immunohistochemistry with Biogenex antibodies showed that 75.8% of previously ER-negative tumors were actually ER positive *(29)*; hence, the use of suboptimal manual hormone receptor assays can result in a false ER-negative result. This would have important

implications in the use of hormone therapy such as tamoxifen as adjuvant therapy.

In other countries in Asia, a similar trend is seen. In Indonesia, ER-positive rates were reported in 51.7% of cases, PgR-positive in 48.2% of cases, and ER or PgR positive in 60.1% of cases *(30)*. In Jordan, 50.8% of cases are ER positive and 57.5% are PgR positive, which is unusual because the ER-positive rate in most studies is higher than the PgR-positive rate *(31)*. In a study of Chinese women in Hong Kong, ER was positive in 53 and 61.6% of the pre- and postmenopausal women, respectively; PgR was positive in 51.5 and 46.2% of women, respectively, and the conclusion was that Chinese patients have lower receptor values and positivity rates than those reported for Caucasians, and that receptor-positive tumors tend to occur in postmenopausal women *(32)*.

Interestingly, the hormone receptor status has been changing over time. Li et al. *(33)* showed that in the United States, ER-positive breast cancers increased from 75.4 to 77.5% from 1992 to 1998, whereas the PgR-positive tumors increased from 65 to 67.7%. The increasing incidence of breast cancer seemed to be in the increase of hormone receptor-positive cancers rather than hormone receptor-negative patients, and it was suggested that hormonal factors may account for this trend. Another study in France, with 11,195 tumors, showed that the percentage of ER-positive tumors rose from 73 to 78% from 1973 to 1992 *(34)*. During this period, the percentage of women older than 50 years had remained the same, and it is unlikely that technical improvements or changes in tumor size, age, or nodal status fully explain this increase; hence, the rising level of ER may reflect a change in breast cancer biology and in hormonal events that influence breast cancer genesis and growth. So far, a similar trend has not been reported in Asia, but perhaps in the next few generations, as the incidence of breast cancer increases, a similar increase in hormone receptor-positive breast cancers may emerge.

The commonest grading system used is the Bloom and Richardson grading system, and in the original study by Bloom and Richardson, 26% of tumors were grade 1, 45% were grade 2, and 34% were grade 3 *(35)*. A recent update by Elston et al. *(36)* reported an 18% grade 1, 27% grade 2, and 45% grade 3. Breast cancers in Asia have often been reported as being higher grade compared with Caucasians. In Indonesia, grade 1 was reported in only 4.1% of breast cancers, grade 2 in 43% and grade 3 in 53% *(37)*. A study from China reported 15.1% grade 1, 45.4% grade 2, and 39.5% grade 3 *(38)*. It is well known that interobserver variation may play a role here, and grading has been shown to be only modestly reproducible *(39)*, although it can be minimized by strict adherence to precise guidelines *(40)*.

Her2-neu overexpression has been reported to occur in 10–40% of breast cancers *(41)*, and there were some reports

that overexpression may be more common in Asia. The incidence of HER2-neu overexpression ranges from 17.5% in Jordan *(31)* to 34.2% in Saudi Arabia *(42)* to 64.2% in Indonesia *(37)*. Interpretation of many studies of HER2, however, is limited by variability in the methods used to detect overexpression and definition of positivity *(43)*, and until there is a large comparative study done between HER2 overexpression in Asians compared with Caucasians, it would be impossible to comment on these isolated reports.

5. Survival of Breast Cancer in Asia

In most countries in the developing world, very little data are available on cancer survival because cancer information systems and even mortality registration are not well established. In addition, there is very little information available on the follow-up of cancer patients *(44)*.

Prognosis from breast cancer is rather good, although globally it still ranks as the leading cost of cancer mortality among women. Five-year survival rates vary from 57% in developing countries to 73% in developed countries *(1)*. Poorer survival in the less affluent developing countries is due to presentation of the disease in later stages, coupled with limited resources; hence, although there is a marked difference in incidence rates, the differences in mortality rates are less marked *(45)*.

The prognosis of Japanese breast cancer patients has steadily improved during the past five decades, and the 10-year overall survival rate after surgery has increased by approximately 5% each decade, from 61% in 1946–1959, 66% in 1960–1969, 72% in 1970–1979, 77% in 1980 –1989, to 83% in 1990–1999. Although several possible causes such as increased life expectancy of Japanese women could have partly contributed to the improved survival of breast cancer patients, it is thought that earlier breast cancer detection and improvements in chemotherapy and hormonal treatments have been the greatest contributors *(12)*. In Korea, the observed 5-year survival rates were 84.1% in 2003 *(8)*. The excellent survival rates in Japan and Korea reflect survival rates in developed countries where early detection programs exist. However, in the rest of Asia, where screening programs for breast cancer are limited, survival rates remain low. In Malaysia, from 1993 to 1997, a 5-year survival rate of 58% was reported *(7)*, whereas the 5-year survival rate in India was reported as 46% in 2002 *(46)*. In Oman, the 5-year survival rate in patients diagnosed from 1996 to 2002 was reported as 64% *(18)*.

Survival from breast cancer depends on 1) early detection, 2) timely and effective treatment, 3) disease factors, and 4) host factors. Because there is no evidence regarding whether there is any difference in disease and host factors between breast cancer in Asia and the rest of the world, the two most important determinants of survival from breast cancer in Asia are early detection and adequacy of treatment.

6. Breast Screening in Asia

At present, the most practical approach to improving the burden of breast cancer is by decreasing the mortality rate through early detection by screening *(47)*. Methods of early detection that have been studied are mammography, breast self-examination (BSE) and clinical breast examination (CBE). Screening by mammography is the most studied modality in the western population; however, there is a paucity of studies in Asia. The randomized controlled trials on mammography screening are the HIP study in 1963 *(48)*, the Swedish two-county study in 1977 *(49)*, the Edinburgh study in 1979 *(50)*, the Malmo study in 1976 *(51)*, the Stockholm trial in 1981 *(52)*, and the Canadian study in 1980 *(53)*, and all these trials except the Canadian and the Malmo trials confirmed an almost 30% mortality after 8 years in the screened group, and this mortality was seen mainly in women 50 years and older. Such studies led to the introduction of population-based mammography screening programs in several developed countries in the 1980s. These programs are expensive, involve labor and training resources, and the success depends on the cooperation of the women themselves.

Singapore embarked on a pilot mammography screening project in the 1995 and reported an uptake rate of only 45%, with a prevalence rate of 2.4 cancers per 1,000 women in the study group compared with 1.3 in the control group *(54)*. A study in Hong Kong concluded that evidence is insufficient to justify population-based mammography screening in Asian populations where the incidence of breast cancer is lower *(55)*.

When mammography screening programs started in several developed countries, breast self-examination was not encouraged. However, in developing countries with no resources for mammography screening, breast self-examination was promoted as the method of early detection of breast cancer, because several previous studies had shown that women who practice BSEs were more likely to be diagnosed with early stages of breast cancer *(56)*. However a controlled trial of 266,064 Chinese women

randomized to BSE or no BSE found that after 10–11 years of follow-up, there was no difference in breast cancer incidence rates or mortality rates between the two groups *(57)*. A recent meta-analysis of trials of BSE training showed that BSE was associated with considerably more women seeking medical advice and having biopsies, but it was not an effective method of reducing breast cancer mortality *(58)*. The term breast self-awareness has now largely replaced breast self-examination, and basically it encompasses education on breast disease and methods of early detection.

Overall, CBE sensitivity ranges from 35 to 56% *(59)*. There are no studies that compare CBE with no screening. A study showed that CBE sensitivity increased with larger tumor size, current users of estrogen and progesterone, and in Asian women and that it decreased with higher body weight and in women in the younger age groups (40–49 years) and >80 years old *(60)*. About three to five women in each 100 women examined will have a false-positive CBE. CBE seems to be a promising means of averting some deaths from breast cancer, whereas BSE seems to have little or no impact on breast cancer mortality *(61)*.

A study on CBE in the Philippines funded by the U.S. Army Medical Research and Material Command *(62)* in 1997 showed that in a single screening round, 151,168 women were offered CBE, of which 92% accepted. Of these women, only 35% completed diagnostic follow-up, whereas 42.4% actively refused further investigation even with home visits, and 22.5% were not traced. The test sensitivity of CBE repeated annually was 53.2%. The actual sensitivity of the program was 25.6%, and the positive predictive value was 1%. Although CBE undertaken by health workers seems to offer a cost-effective approach to reducing mortality, the sensitivity of the screening program in the real context was low. Moreover, in this relatively well-educated population, cultural and logistic barriers to seeking diagnosis and treatment persist and need to be addressed before any screening program is introduced.

Many Asian women are not aware of the importance of regular screening. Cultural attitudes toward breast cancer screening tests, modesty, and lack of encouragement by family members and physicians are the major inhibitors to women's participation in breast cancer screening. Health education using media and community health programs to create awareness of the advantages of earlier presentation and diagnosis of breast cancer in Asian women can motivate participation in breast cancer screening programs *(63)*. A breast screening program also must be developed together with development of adequate facilities for investigation and treatment of the expected increase in the number of cases of breast cancer detected.

7. Treatment of Breast Cancer in Asia

Access to treatment is an important determinant of survival from breast cancer. Most countries in Asia are low- and middle-income countries, and in such countries there are four key problems in the health care systems. First, public facilities, which provide most secondary and tertiary health care in most countries, offer poor-quality services. Second, providers cannot be enticed to rural and urban marginal areas, leaving large segments of the population without adequate access to health care. Third, the composition of health services offered and consumed is suboptimal. And fourth, coordination in the delivery of care, including referrals, second opinions, and teamwork, is inadequate *(64)*.

Evidence-based guidelines outlining optimal approaches to breast cancer detection, diagnosis, and treatment have been well developed and disseminated in several high-resource countries *(65,66)*. Most guidelines define optimal practice that may be inappropriate to apply in most countries in Asia for numerous reasons, including inadequate numbers of trained health care providers; inadequate diagnostic and treatment infrastructure, such as operating rooms, radiation therapy equipment, pathology, pharmacy, infusion centers, and microbiology laboratories; lack of drugs; lack of radiographic film; inadequate transportation systems; and cultural, societal, or religious barriers to women accessing the health care system. Thus, in a country with limited resources, many barriers exist between the average patient and the level of care dictated by guidelines applicable to high-resource settings.

Hence, there is a need to develop clinical practice guidelines oriented toward countries with limited financial resources *(67)*. It was for this purpose that the Breast Health Global Initiative (BHGI) was established in 2002. Co-sponsored by Fred Hutchinson, the BHGI is a program that strives to develop, implement, and study evidence-based, economically feasible, and culturally appropriate guidelines that can be used in low- and middle-income countries, with the aim of improving breast outcomes. The stepwise, systematic approach to health care improvement outlined by the 2005 BGHI panels involved a tiered system of resource allotment defined using four levels—basic, limited, enhanced, and maximal—based on the contribution of each resource toward improving clinical outcomes *(68–72)*.

- Basic level—Core resources or fundamental services absolutely necessary for any breast health care system to function. By definition, a health care system lacking any basic level resource would be unable to provide breast cancer care to its patient population. Basic level services are typically applied in a single clinical interaction.

- Limited level—Second-tier resources or services that produce major improvements in outcome such as increased survival, but which are attainable with limited financial means and modest infrastructure. Limited-level services may involve single or multiple clinical interactions.
- Enhanced level—Third-tier resources or services that are optional but important. Enhanced level resources may produce minor improvements in outcome but increase the number and quality of therapeutic options and patient choice.
- Maximal level—High-level resources or services that may be used in some high resource countries and/or may be recommended by breast care guidelines that assume unlimited resources, but that should be considered a lower priority than those in the basic, limited, or enhanced categories on the basis of extreme cost and/or impracticality for broad use in a resource-limited environment. To be useful, maximal-level resources typically depend on the existence and functionality of all lower level resources.

Improving breast health care in Asia requires collaboration of multiple sectors, both public and private. Ultimately, it will be political which will be the decisive factor, and this is the area where advocacy can play a role in galvanizing the politics that will meet the challenge of improving breast care so that women with breast cancer in limited resource countries are able to access the care they need to survive breast cancer, the commonest cancer in women in most parts of Asia.

References

1. Parkin DM, Whelan SL, Ferlay J, Storm H (2005) Cancer incidence in five continents, volumes I–VIII. IARC CancerBase no. 6. IARC Press, Lyon, France (http://www-dep.iarc.fr).
2. Seow A, Duffy SW, McGee MA, Lee J, Lee HP. (1996). Breast cancer in Singapore: trends in incidence 1965–1992. Int J Epidemiol **25**, 40–5.
3. Yang BH, Parkin DM, Cai L, Zhang ZF. (2004). Cancer burden and trends in the Asian Pacific Rim Region. Asian Pac. J Cancer Prev **5**, 96–117.
4. Ziegler RG, Hoover RN, Pike MC, et al. (1993). Migration patterns and breast cancer risk in Asian-American women. J Natl Cancer Inst **85**, 1819–27.
5. Bray F, McCarron P, Parkin DM. (2004). The changing global patterns of female breast cancer incidence and mortality. Breast Cancer Res 6, 229–39.
6. Lim GCC, Halimah Y (eds). (2004). Cancer incidence in Malaysia 2003. National Cancer Registry, Kuala Lumpur, Malaysia.
7. Yip CH, Taib NA, Mohamed I. (2006). Epidemiology of breast cancer in Malaysia. Asian Pac J Cancer Prev **7**, 369–74.
8. Son BH, Kwak BS, Kim JK, et al. (2006). Changing patterns in the clinical characteristics of Korean patients with breast cancer during the last 15 years. Arch Surg **141**, 155–60.
9. Henderson BE, Ross R, Bernstein L. (1988). Estrogens as a cause of human cancer: the Richard and Hinda Rosenthal Foundation Award Lecture. Cancer Res **48**, 246–53.
10. Moolgavkar SH, Stevens RG, Lee JA. (1979). Effect of age on incidence of breast

cancer in females. J Natl Cancer Inst **62**, 493–501.

11. Minami Y, Tsubono Y, Nishima Y, Ohuchi N, Shibuya D, Hisamichi S. (2004). The increase in female breast cancer incidence in Japan: emergence of a birth cohort effect. Int J Cancer **108**, 901–6.
12. Yoshimoto M, Tada K, Hori H, et al. (2004). Improvement in the prognosis of Japanese breast cancer patients from 1946–2001: an institutional review. Jpn J Clin Oncol **34**, 457–62.
13. Hisham AN, Yip CH. (2003). Spectrum of breast cancer in Malaysian women: an overview. World J Surg **27**, 921–23.
14. Aggarwal G, Pradeep PV, Aggarwal V, Yip CH, Cheung PS. (2007). Spectrum of breast cancer in Asian women. World J Surg **31**, 1031–40.
15. Remenick L. (2006). The challenge of early breast cancer detection among immigrant and minority women in multicultural societies. Breast J **12**, S103–S110.
16. Parsa P, Kandiah M, Abdul Rahman H, Mohd Zulkefli NA. (2006). Barriers for breast cancer screening among Asian women: a mini literature review. Asian Pac J Cancer Prev **7**, 509–14.
17. Hebert JR, Ghumare SS, Gupta PC. (2006). Stage at diagnosis and relative differences in breast and prostate cancer incidence in India: comparison with the United States. Asian Pac J Cancer Prev **7**, 547–55.
18. Al-Moundhari M, Al-Bahrani B, Pervez J, et al. (2004). The outcome of treatment of breast cancer in a developing country–Oman. Breast **13**, 139–45.
19. Vahdaninia1 M, Montazeri A. (2004). Breast cancer in Iran: a survival analysis. Asian Pac J Cancer Prev **5**, 223–25.
20. Ahn SH, Yoo KY. (2006). The Korean Breast Cancer Society. Chronological changes of clinical characteristics in 31,115 new breast cancer patients among Koreans during 1996–2004. Breast Cancer Res Treat **99**, 209–14.
21. 21.Tan EY, Wong HB, Ang BK, Chan MY. (2005). Locally advanced and metastatic breast cancer in a tertiary hospital. Ann Acad Med Singap **34**, 595–601.
22. 22..Klonoff-Cohen HS, Schaffroth LB, Edelstein SL, Molgaard C, Saltzstein SL. (1998). Breast cancer histology in Caucasians, African Americans, Hispanics, Asians and PacificIslanders. Ethn Health **3**, 189–98.
23. Sakamoto G, Sugano H. (1991). Pathology of breast cancer: present and prospect in Japan. Breast Cancer Res Treat **18**, S581–S583.
24. Harirchi I, Karbakhsh M, Kashefi A, Momtahen AJ. (2004). Breast cancer in Iran: results of a multicentre study. Asia Pac J Cancer Prev **5**, 24–7.
25. CH Yip, LM Looi. (1996). Breast cancer in Malaysian women–pathological features and treatment. Asian J Surg **19**, 112.
26. Gapstur SM, Dupuis J, Gann P, Collila S, Winchester DP. (1996). Hormone receptor status of breast tumors in black, Hispanic, and non-Hispanic white women. An analysis of 13,239 cases. Cancer **77**, 1465–71.
27. Li CI Malone KE, Daling JR. (2002). Differences in breast cancer hormone receptor status and histology by race and ethnicity among women 50 years of age and older. Cancer Epidemiol Biomarkers Prev **11**, 601–7.
28. Desai SB, Moonim MT, Gill AK, Punia RS, Naresh KN, Chinoy RF. (2000) Hormone receptor status of breast cancer in India: a study of 798 tumours. Breast **9**, 267–70.
29. 29.Navani S, Bhaduri AS. (2005). High incidence of oestrogen receptor negative progesterone receptor positive phenotype in Indian breast cancer: fact or fiction? Indian J Pathol Microbiol **48**, 199–201.
30. Aryandono T, Harijadi, Soeripto. (2006). Hormone receptor status of operable breast cancer patients in Indonesia: correlation with other prognostic factors and survival. Asian Pac J Cancer Prev **7**, 321–4.
31. 31.Sughayer MA, Al-Khawaja MM, Massarweh S, Al-Masri M. (2006). Prevalence of hormone receptors and HE2/neu in breast cancer cases in Jordan. Pathol Oncol Res **12**, 83–6.
32. Chow LW, Ho P. (2000). Hormonal receptor determination of 1,052 Chinese breast cancers. J Surg Oncol **75**, 172–5.
33. Li CI, Daling JR, Malone KE. (2003). Incidence of invasive breast cancer by hormone receptor status from 1992 to 1998. J Clin Oncol **21**, 28–34.
34. Pujol P, Hilsenbeck SG, Chamness GC, Elledge RM. (1994). Rising levels of estrogen receptor in breast cancer over 2 decades. Cancer **74**, 1601–6.
35. Bloom HJG, Richardson WW. (1957) Histological grading and prognosis in breast cancer. Br J Cancer **11**, 359–77.
36. Elston CW, Ellis IO. (2002). Pathological prognostic factors in breast cancer. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology **41**, 154–61.
37. Aryandono T, Harijadi, Soeripto. (2006). Survival from operable breast cancer: prognostic

factors in Yogyakarta, Indonesia. Asian Pac J Cancer Prev **6**, 455–9.

38. Zhang T, Tu X, Xu W. (1998). A study of prognostic factors in breast cancer: histological grading. Zhong Hua Bing Li Xue Za Zhi **27**, 405–7
39. Meyer JS, Alvarez C, Milikowski C, et al. (2005). Breast carcinoma malignancy grading by Bloom–Richardson system vs proliferation index: reproducibility of grade and advantages of proliferation index. Mod Pathol **18**, 1067–78.
40 40 Sikka M, Agarwal S, Bhatia A. (1999). Interobserver agreement of the Nottingham histologic grading scheme for infiltrating duct carcinoma breast. Indian J Cancer **36**, 149–53.
41. Menard S, Taqliabue E, Campiqlio M, Pupa SM. (2000). Role of HER2 gene overepression in breast carcinoma. J Cell Physiol **182**, 150–62.
42. Homael-Shandiz F, Ghavam-Nassiri MR, Sharif N, et al. (2006). Evaluation of the relationship between human epidermal growth factor receptor-2/neu (c-erbB-2) amplification and pathologic grading in patients with breast cancer. Saudi Med J **27**, 1810–4.
43. Cianfrocca M, Goldstein LJ. (2004). Prognostic and predictive factors in early stage breast cancer. Oncologist **9**, 606–16.
44. Sankaranarayanan R, Swaminathan R, Black RJ. (1996). Global variations in cancer survival. Cancer **78**, 2461–4.
45. Pal SK, Mittal B. (2004). Fight against cancer in countries with limited resources: the postgenomic era scenario. Asian Pac J Cancer Prev **5**, 328–33.
46. Parkin DM, Bray F, Ferlay J, Pisani P. (2005). Global cancer statistics, 2002. CA Cancer J Clin **55**, 74–108.
47. Moore MA, Kazuo T, Pham Hoang Anh, et al. (2003). Grand challenges in global health and the practical prevention program? Asian focus on cancer prevention in females of the developing world. Asian Pac J Cancer Prev **4**, 153–65.
48. Shapiro S, Venet W, Strax P, et al. (1988). Periodic screening for breast cancer: the Health Insurance Plan project and its sequelae, 1963–1986. Johns Hopkins University Press, Baltimore, MD.
49. Nystrom Lrutqvist LE, Wall S, et al. (1993). Breast screening with mammography: overview of the Swedish Randomized Trials. Lancet **17**, 973–8.
50. Roberts MM, Alexander FE, Anderson TJ, et al. (1990) Edinburgh trial of screening for breast cancer: mortality at 7 years. Lancet **335**, 241–6.
51. Andersson I, Aspergren K, Janzon L, et al. (1988). Mammographic screening and mortality from breast cancer: the Malmo mammographic screening trial. BMJ **297**, 943–8.
52. Frisell J, Eklund G, Hellstrom L, et al. (1991). Randomized study of mammography screening-preliminary report on mortality in the Stockholm trial. Breast Cancer Res Treat **18**, 49–56.
53. Miller AB, To T, Baines CJ, Wall C. (2002). The Canadian National Breast Screening Study 1: breast cancer mortality after 11 to 16 years of follow-up. A randomized screening trial of mammography in women aged 40–49 years. Ann Intern Med **137**, 305–12.
54. Ng EH, Ng FH, Tan PH, Low SC, Chiang G, et al. (1998). Results of intermediate measures from a population based randomized trial of mammographic screening-prevalence and detection of breast carcinoma among Asian women. Cancer **82**, 1521–8.
55. Leung GM, Lam TH, Thach TQ, Hedley AJ. (2002). Will screening mammography in the east do more harm than good? Am J Public Health **92**, 1841–6.
56. Hill D, White V, Jolley D, Mapperson K. (1988). Self examination of the breast: is it beneficial? A meta-analysis of studies investigating breast self-examination and extent of disease in patients with breast cancer. BMJ **297**, 271–5.
57. Thomas DB, Gao DL, Ray RM, Wang WW, Allison CJ, et al. (2002). Randomized trial of breast self-examination in Shanghai: final results. J Natl Cancer Inst **94**, 1445–57.
58. Hackshaw AK, Paul EA. (2003). Breast Self examination and death from breast cancer: a meta-analysis. Br J Cancer **88**, 1047–53.
59. Bancej C, Decker K, Chiarelli A, Harrison M, Turner D, Brisson J. (2003). Contribution of CBE to mammography screening in the early detection of breast cancer. J Med Screen **10**, 16–21.
60. Oestreicher N, White E, Lehman CD, Mandelson MT, Porter PL, Taplin SH. (2002). Predictors of sensitivity of CBE. Breast Cancer Res Treat **76**, 73–81.
61. Weiss NS. (2003). Breast cancer mortality in relation to CBE and BSE. Breast J **9**, S86–S89
62. Pisani P, Parkin DM, Ngelangel C, Esteban D, et al. (2006). Outcome of screening by clinical examination of the breast in a trial in the Philippines. Int J Cancer **118**, 149–54.

63. Parisa P, Kandiah M, Abdul Rahman H, Mohd Zulkefli NA. (2006). Barriers to breast cancer screening among Asian women: a mini literature review. Asian Pac J Cancer Prev **7**, 509–514.
64. Varun G. (2001). Are incentives everything? Payment mechanisms for health care providers in developing countries. World Bank Policy Research Working Paper No. 2624. (http://ssrn.com/abstract=632692).
65. Goldhirsch A, Glick JH, Gelber RD, Coates AS, Thurlimann B, Senn HJ. (2005). Meeting highlights: international expert consensus on the primary therapy of early breast cancer. Ann Oncol **16**, 1569–83.
66. Carlson RW, Anderson BO, Burstein HJ, Cox CE, Edge SB, Farrar WB, et al. (2005) Breast cancer. J Natl Compr Canc Netw **3**, 238–89.
67. World Health Organization. (2002). Executive summary of the national cancer control programmes: policies and managerial guidelines. World Health Organization, Geneva, Switzerland.
68. Smith RA, Caleffi M, Albert US, Chen TH, Duffy SW, Franceschi D, et al. (2006). Breast cancer in limited-resource countries: early detection and access to care. Breast J **12**, S16–S26.
69. Shyyan R, Masood S, Badwe RA, Errico KM, Liberman L, Ozmen V, et al. (2006). Breast cancer in limited-resource countries: diagnosis and pathology. Breast J **12**, S27–S37.
70. Eniu A, Carlson RW, Aziz Z, Bines J, Hortobagyi GN, Bese NS, et al. (2006). Breast cancer in limited-resource countries: treatment and allocation of resources. Breast J **12**, S38–S53.
71. Anderson BO, Yip CH, Ramsey SD, Bengoa R, Braun S, Fitch M, et al. (2006). Breast cancer in limited-resource countries: health care systems and public policy. Breast J **12**, S54–S69.
72. Anderson BO, Shyyan R, Eniu A, Smith RA, Yip CH, Bese NS, et al. (2006). Breast cancer in limited-resource countries: an overview of the Breast Health Global Initiative 2005 guidelines. Breast J **12**, S3–S15.

Chapter 4

Cancer Epidemiology in the United States: Racial, Social, and Economic Factors

Dana Sloane

Summary

It is widely accepted that there is a differential burden of cancer in certain populations, including racial/ethnic minorities, the medically underserved, and older adults. Differences in survival, stage at diagnosis, and risk of death have been identified in these populations for cancers of the lung, colon and rectum, prostate, and female breast. The factors that drive these disparities are not uniformly understood. Addressing the unique issue of racial differences in cancer epidemiology necessitates a discussion of the definitions of "race" and "ethnicity," and an analysis of the validity of these concepts within the context of scientific study. Poor cancer-related health outcomes in groups of low socioeconomic status highlight issues of access to care and preventive care use. There is a scant amount of data on cancer in the elderly, and the special considerations that this group faces. A unique challenge facing cancer epidemiologists is suboptimal recruitment of members of these groups into clinical studies, which precludes a robust understanding of the existing disparities. It is critical to appreciate the overlap that exists between these populations, because this may complicate data interpretation. Legislative efforts that have, in part, been driven by the National Center on Minority Health and Health Disparities and by the Department of Health and Human Services, will continue to play an instrumental role in the identification and resolution of cancer disparities in these groups.

Key Words: Cancer, disparity, race, ethnicity, socioeconomic status, elderly, cancer survival, stage at diagnosis, cancer mortality.

1. Introduction

A fundamental requirement of cancer epidemiology is to describe—and critically distinguish—the populations that are affected by this disease. It is widely understood that there is a differential burden of cancer in certain populations, including racial/ethnic minorities, the medically underserved, and older adults. Thus, if cancer epidemiology is to be a true cornerstone of public health, it is critical to understand the nature of these

Mukesh Verma (ed.), *Methods in Molecular Biology, Cancer Epidemiology, vol. 471*
© 2009 Humana Press, a part of Springer Science + Business Media, Totowa, NJ
Book doi: 10.1007/978-1-59745-416-2

differences, *why* these populations are disproportionately affected, and *how* these disparities can best be addressed.

The topic of racial/ethnic, cultural, and social factors, as they pertain to cancer epidemiology, is, admittedly, quite broad. The breadth and depth of relevant information would be enough to fill several volumes. For the sake of brevity, and to more carefully review these issues, our discussion here is limited. First, although there are dozens of types of known cancers, the four dominant cancers in the United States—lung, colorectal, male prostate, and female breast—are featured here. Similarly, the discussion is limited to certain subgroups of the population. There are innumerable permutations, within the total U.S. population, of racial/ethnic, cultural, and social groups. Persons may be categorized by gender, age, race, ethnicity, education level, income, geographic location, disability, social class, religious affiliation, or sexual orientation. Beyond that, each of these subpopulations has its own inherent heterogeneity. These categories also often overlap. The sheer scope of these issues cannot be confined to this chapter. However, by focusing on select large groups—certain racial/ethnic minorities, the economically disadvantaged, and older adults in these populations—this chapter seeks to create a platform for more robust consideration of cancer-related health issues.

2. Defining Unique Populations

One of the aforementioned populations—racial/ethnic minorities—bears special consideration before an in-depth examination can be pursued. The scientific validity of "race" has been an inflammatory issue for decades. It has been proposed that race is, fundamentally, a social construct that has no anthropologic or scientific basis. There is both truth and fallacy in this argument, and it is important to address both, within the context of cancer epidemiology.

Race and ethnicity are critical entities in the United States, inextricably woven into the fabric of this country's social history and culture. As stated by Whaley, "culture is the context in which epidemiological research is performed" *(1)*. It is therefore important to recognize the crucial roles that race and ethnicity have played in the public health history of the United States.

What is race? The application of the term, as it is currently used, has its foundation in the late eighteenth century, when it was used to reference social categories in the United States. (i.e., Indians, Blacks, Whites) *(2)*. Over the past century, this term has evolved into a two-pronged perception of race: one that emphasizes genetic variation, the other that takes into account

perceived differences in phenotypical features and behavior *(3)*. The latter understanding of race is certainly the most widely used. It has been suggested that these two definitions of race are mutually exclusive and that the commonly used definition of race does not, in fact, capture true genotypic heterogeneity *(4)*. The case also could be made that classifications based on race are too limiting, and that they do not capture multidimensional differences based on cultural, environmental, and social factors *(5)*. Thus, some have advocated the elimination of race as a variable in epidemiologic research *(6,7)*. This suggestion was translated into political action in 2003, when California's Proposition 54 gained national attention. In summary, this legislation sought to eliminate the collection and use of data on race, ethnicity, and national origin by all public agencies in the state, and by any group that received state funding; this would have included elimination of racial/ethnic classification of medical data. The proposition was ultimately defeated; however, its very existence illustrates the controversy regarding the role of race and ethnicity in science and public health.

The counterpoint to this position, which would support the inclusion of race in scientific study, contends that racial categories have both contributed to and perpetuated social, economic, and political disadvantages that continue to impact racial differences in health outcomes *(4)*. If this effect is real, then careful assessment of race as it pertains to risk factor exposure, biological precursors, and cancer prevention, is valid.

Further complicating this debate is the evolving perception of racial classification in the United States. The 2000 census was the first to offer respondents the opportunity to select more than one race. Interestingly, 2.4% of the U.S. population identified with two or more races. According to the analysis by the Social Science Data Analysis Network, these new classification options result in >120 racial and ethnic categories *(8)*. Beyond the obvious issues of possible reporting bias, consider the enormous implications this will have on large-scale data collection and analysis in epidemiologic research. One must wonder if science will—or should—be able to keep pace with society's rapidly changing understanding of race.

Should there be a different classification system in epidemiologic research? A proposed alternative is the use of ethnicity. The definitions for ethnicity are, not surprisingly, diverse; most emphasize unifying cultural traditions, ancestry, national origin, history, or religion. In 1999, the Institute of Medicine (IOM) published *The Unequal Burden of Cancer*, in which it advocated "an emphasis on ethnic groups rather than on race in the [National Institutes of Health] cancer surveillance" *(9)*. Current categories of ethnicity defined by the U.S. Office of Management and Budget (OMB) are limited to Hispanic or Latino (which may

include any race), versus neither; racial categories are delineated as Asian, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Black or African American, and White. The OMB actually states that its categories "represent a *social-political construct* designed for collecting data on the race and ethnicity of broad population groups" [emphasis added]. In *The Unequal Burden of Cancer*, the authors advocate that ethnicity should be used to classify the population, because race does not imply any natural biological differences, and ethnicity implies cultural, behavioral and environmental factors that … affect the risk of disease *(10)*.

Although, in theory, this proposal may be well founded, one must wonder whether it is truly practical, because race in the United States has practically become an entity unto itself. Oppenheimer made an interesting point in his 2001 argument, speculating whether "race—an ideologically powerful category during most of U.S. history … can so easily be expunged and whether something significant may not be lost by excising race as an analytic category." This perspective is intriguing, because it recognizes the significant impact of race in this society (and, by default, on public health), without implying biological significance that may or may not be valid.

A final note on this issue of racial/ethnic classification is that of data collection. As mentioned, self-reporting of racial/ethnic identity in this era of diversity may not be easily translated into efficient data gathering. However, what would be a viable alternative? If data collectors or investigators were made responsible for categorizing people by race, what criteria would be followed? There is the obvious potential for misclassification, as illustrated by Frost and colleagues. The authors identified racial misclassification of American Indians in the Surveillance, Epidemiology and End Results (SEER) registry, which contributed, in part, to perceived lower cancer incidence in this population *(11)*.

Racial/ethnic classifications have historically afforded—and withheld—economic and political opportunities, creating a socioeconomic hierarchy in the United States, with far-reaching implications. The notion that race is not, to some extent, a predictor of income, insurance, and access to care, is quickly disputed by a review of the nation's vital statistics. Interestingly, racial and ethnic disparities in health care tend to persist even after adjustment for socioeconomic differences, suggesting that there are other, less tangible factors that interact between race and health outcomes.

Clearly, use of race and ethnicity, particularly within the context of population-centered research, is fraught with complication. Whether o not race has a legitimate place in scientific discussions, race does permeate many aspects of our society, and as such, it warrants careful examination in any discussion of cancer epidemiology. For the purpose of this discussion, race/

ethnicity is used as conjoined terms. However, as stated in **Subheading 1**, there are other subgroups of the population that are disproportionately burdened by cancer, including the medically underserved.

Although the term "medically underserved" is not burdened by the same controversy as race and ethnicity, ambiguity remains. In *The Unequal Burden of Cancer*, the IOM describes the medically underserved population as including "underinsured or uninsured people; those with low levels of education; rural and inner-city populations; unemployed people; or those with low socioeconomic status (SES)." Thus, this is a broad category, and it may take into account any combination of these groups.

3. Defining Disparity

It has been well established that differences in rates of disease and in health outcomes exist between people of different racial/ethnic origin, socioeconomic status, and area of residence. The 2004 report of the U.S. Department of Health and Human Services' (HHS) Trans-HHS Cancer Health Disparities Progress Review Group *(12)* stated that "minority and underserved populations … are significantly more likely to: be diagnosed with and die from preventable cancers; be diagnosed with late-stage disease for cancers detectable through screening in the early stage; receive either no treatment or treatment that does not meet currently accepted standards of care; [and] die of cancers that are generally curable."

A critical component to a discussion of cancer epidemiology in diverse populations is that of health disparity. But, as pointed out in a recent discussion by Thomson and colleagues, health *disparities* are not merely *differences* in health *(13)*. What, then, is the appropriate use of the term "disparity"? A clear definition of this word is crucial, to thoroughly understand the epidemiology of cancer in these populations, and to more appropriately design and conduct epidemiologic research. However, there is controversy over the appropriate definition *(14)*. To paraphrase language used by the World Health Organization, the crux of the debate on disparity centers on health "inequities" versus "inequalities" *(15–17)*. Although the former term implies a difference that is inequitable or unjust, the latter term does not necessarily denote injustice. Whitehead defines health inequities as "differences in health which are not only unnecessary and avoidable, but in addition, are considered unfair and unjust" *(18)*. This discussion further details several determinants of health disparity that could be categorized as *unjust*, including health-damaging

behaviors in the context of restricted lifestyle options, exposure to unhealthy living and/or working conditions, and inadequate access to essential health care or other basic services. In the context of epidemiology and public health, disparity has been used interchangeably with inequity, that is, assuming the implication of injustice. Does disparity necessarily imply inequity? This is a complicated question, involving judgment of what is avoidable and fair, for which there is no simple answer.

An interesting concept within this discussion of disparity and inequity is the language that will be used in the upcoming years—by epidemiologists, those in public health, and society at-large. Specifically, some ethnic "minorities," in particular, the Hispanic population, will not be in the minority for much longer. Data from the 2000 census project that Hispanics/Latinos will represent nearly 25 percent of the U.S. population by the year 2050 *(19)*. Thus, the very terminology used in these discussions of inequity and disparity will likely need to be re-examined, in the near future.

This controversy is not purely academic. Defining disparity becomes crucial when one considers the need to accurately measure disparity, and to then apply these assessments to health funding and policy. Carter-Pokras and Baquet have made the argument that policy makers must look beyond issues of difference and inequality, to consider which of these differences are truly inequitable, and thus, in need of reform *(17)*. This important distinction may facilitate prioritization of public health research efforts. Of course, the judgment of inequity is subjective, and these decisions will inevitably require a great deal of debate at a multidisciplinary level.

4. Burden of Disease on Unique Populations: Racial/Ethnic Minorities, Poor/Uninsured, Older Adults

There is certainly an association between race/ethnicity and cancer. In fact, distinctive cancer patterns have been identified for various racial/ethnic groups *(20,21)*. It also has been well established that racial/ethnic minorities, compared with their White majority counterparts, are more likely to present at a later stage of cancer, and less likely to have a favorable outcome from cancer. The disproportionate burden of cancer in these populations has important implications, particularly with regard to matters of disease control, access to care (including screening and treatment), and public health research.

There was an overall decline in cancer death rates in the United States, first noticed in the early 1990s, which continued into 2003 *(22,23)*. These improvements have been attributed to advancements made in prevention efforts, including exposure reduction, enhanced screening measures, more effective cancer treatment, and more

widespread access to care. Although cancer survival rates are also improving overall, disturbing disparities remain.

Despite having a lower cancer mortality rate than Whites in 1950, African Americans had a 30% *higher* cancer-related mortality rate in 2000 *(24)*. As of the same year, African Americans were 34% more likely to die of cancer than were Whites; they were more than twice as likely to die of cancer as Asian or Pacific Islanders, American Indians, and Hispanics *(25)*. A gender influence exists as well: African American males have the highest death rates of colon and rectum, lung, and prostate cancers. In addition, a recent population-based study demonstrated that racial/ethnic differences in survival rates and relative risk of cancer-related death have persisted for all cancers (including lung, colorectal, breast and prostate *[26]*). Specifically, American Indians and Alaskan Natives have the lowest survival rates for cancers of the breast, lung, prostate, and for all cancers combined. Both Hispanic and non-Hispanic Whites and Asian Americans had the best survival rates for all cancer sites. These racial/ethnic differences persisted after controlling for age and stage. Compared with non-Hispanic Whites, the relative risk of death for every racial/ethnic minority group—with the exception of Asian Americans—was higher for each of the four cancer types and for all cancers combined (adjusted risk ration [RR] = 1.1-2.0).

It is intriguing to note that not *all* racial/ethnic minorities have higher cancer-related mortality, Asian Americans serving as a case-in-point. One contributing factor may be that members of these populations choose to die in their native countries, leading to cancer death underreporting in the United States *(27)*. However, underreporting may only be part of the explanation. Are there less obvious reasons why certain racial/ethnic groups have better outcomes? Because much attention is paid to the cause for higher incidence and mortality in certain minority groups, it may be equally interesting to further explore the reasons for *lower* incidence and mortality in others. It is conceivable that there are undiscovered clues about the effect of diet and other health behaviors within these subgroups that could benefit the population at large.

The study by Clegg and colleagues *(26)*, which was the first to examine the risk of cancer survival and death across the six major racial/ethnic groups in the United States, also demonstrated differences in stage at diagnosis. From 1988 to 1997, American Indians and Alaskan Natives had the highest percentages of distant stage cancer of the lung, prostate and breast; together with the latter groups, African Americans shared the highest percentages of distant-stage colorectal cancer. As described in a more recent national report on the status of cancer in the United States, Hispanics and African Americans are less likely to be diagnosed with localized disease (among cancers of the lung, colon and rectum, prostate, and female breast) than Whites *(23)*. **Table 4.1** illustrates that these minority groups are more likely to present

Table 4.1
Stage distributions[a] of cases for selected common cancer Sites[b] and age groups[c] by race/ethnicity[d] and sex in selected areas[e] of the United States, 2001–2003

Cancer site	Stage	Non-Hispanic White	Hispanic	Non-Hispanic Black
Males				
Colon and rectum				
	Localized	41.8	36.5	36.7
	Regional	40.6	43.4	38.9
	Distant	17.6	20.1	24.4
Lung and bronchus				
	Localized	19	14.9	15.4
	Regional	28	24.4	28
	Distant	53	60.7	56.6
Prostate				
	Localized	86.5	83.7	83.9
	Regional	9.8	10.9	9.3
	Distant	3.7	5.4	6.8
Females				
Breast				
	Localized	66.2	58.3	54.9
	Regional	29.4	36.6	37.4
	Distant	4.4	5.2	7.8
Colon and rectum				
	Localized	40.1	38.4	36.3
	Regional	42.7	42.7	41.8
	Distant	17.2	18.8	22
Lung and bronchus				
	Localized	22.4	19	17.8
	Regional	27.1	25.3	27.8
	Distant	50.5	55.7	54.4

Adapted from: Howe HL, Wu X, Ries LA. Annual report to the nation on the status of cancer, 1975-2003, featuring cancer among U.S. Hispanic/Latino populations.
Source: SEER and NPCR areas reported b the North American Association of Central Cancer Registries as meeting high quality standards for 1999–2003.[a] The 2000 SEER summary staging system. [b] Common cancers for age groups selected based on available tests for diagnosis in early stages of disease progression. [c] NHIA-derived Hispanic origin. [d] Colon and rectum: age >50; lung and bronchus: age>20; prostate: age>50; and breast: age>40.[e] The data from 36 cancer registries (Alabama, Alaska, California, Colorado, Connecticut, Delaware, District of Columbia, Florida, Georgia, Idaho, Illinois, Inidana, Iowa, Kentucky, Louisiana, Maine, Massachusetts, Michigan, Minnesota, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New York, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, Texas, Utah, Washington, West Virginia, and Wisconsin) were included covering 80% of United States, 89% of Hispanic, 79% of non-Hispanic Whites, and 75% of non-Hispanic Black populations.

with regional and distant disease. These patterns of more advanced stage at diagnosis suggest that the lower survival rates for certain minority groups, particularly African Americans, may be explained by lack of early detection *(28)*. This possibility provides special public health challenges, with regard to enhancing targeted screening and preventive efforts.

A common putative explanation for the racial/ethnic disparities in survival, stage at diagnosis, and risk of death is that differences exist in access to, and use of, preventative health measures. This concept opens the door to countless possible risk factors, including economic, behavioral, cultural, genetic, and even occupational factors, which may influence these differences. When considering these issues, it is important to note that race and ethnicity often overlap with socioeconomic status. According to 2001 data from the National Health Interview Surveys, 40.1% of Hispanics and 22.8% of Blacks between the ages of 18 and 64 years were without health insurance coverage, compared with only 13.5% of Whites *(29)*. Obviously, health insurance is closely associated with income, educational status, and overall economic well-being *(30)*. Data from the 2000 census reflect this relationship, indicating that racial/ethnic minorities have higher poverty rates than those in the White majority (**Fig. 4.1**). Data from the past 10 years also have demonstrated lower educational levels among minorities, including non-White Hispanics *(19)*, which is closely

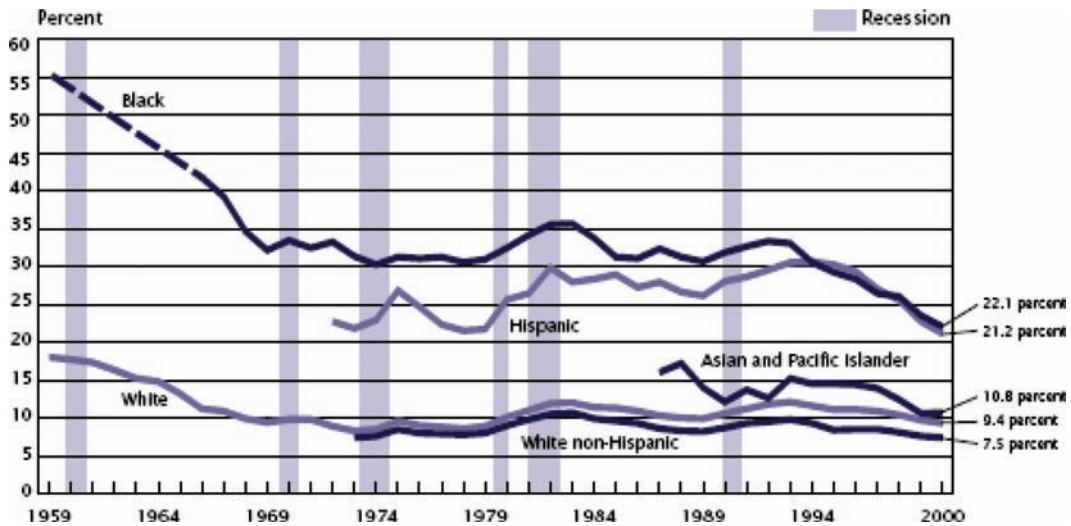


Fig. 4.1. Poverty rates by race and Hispanic origin: 1959–2000. Adapted from Dalaker J. Poverty in the United States: 2000. U.S. Census Bureau, U.S. Department of Commerce Economics and Statistic Administration. Issued September 2001. Note: The data points represent the midpoints of the respective years. The latest recession began in July 1990 and ended in March 1991. Data for Blacks are not available from 1960 to 1965. Data for the other race and Hispanic origin groups are shown from the first year available. Hispanics may be of any race.
Source: U.S. Census Bureau, Current Population Survey, March 1960–2001.

correlated with lower economic well-being and health status. Low educational level also is related to suboptimal use of screening options, as is discussed in **Subheading 5**.

Although it may be tempting to summarily attribute the racial/ethnic variations in cancer-related survival and mortality to socioeconomic status, there are many examples of studies that demonstrate persistent – albeit attenuated – racial/ethnic differences in stage at diagnosis, even after adjusting for socioeconomic factors *(31–34)*. This would suggest that although socioeconomic status plays an integral role in the relationship between racial/ethnic background and cancer disparity, there are other contributing factors, which may be less apparent. These factors may include residential segregation, discrimination, and/or poor health practices *(35)*.

Although disparities in survival, stage at diagnosis and mortality cannot entirely be explained by the association between race/ethnicity and socioeconomic status, there is clearly an important relationship between SES and health outcomes. Thus, it is relevant to review the disproportionate burden of cancer in populations with low socioeconomic status.

According to data from the SEER Program, area socioeconomic gradients were identified in all-cancer mortality for both men and women between 1975 and 1999 *(20)*. Specifically, total cancer mortality among men was 13% greater in high poverty areas than in low poverty areas. (Of note, SEER defines high poverty areas as those census tracts with poverty rates greater than or equal to 20%.) The gradient in total mortality was not as pronounced among women in high poverty areas (3% higher). Socioeconomic gradients were most prominent for Hispanics: total cancer mortality was 45% higher for Hispanic men and 35% higher for Hispanic women, in high poverty areas than in low poverty areas. Interestingly, this gradient was reversed in Asian/Pacific Islander women. Socioeconomic gradients followed a similar pattern of higher incidence and mortality rates in high poverty areas for lung cancer, and of higher mortality rates for colorectal, prostate, and breast cancer.

These mortality data mirror that of survival: Both men and women in high poverty areas had lower 5-year survival rates than their counterparts in low poverty areas. Specifically, among men in high poverty areas who were diagnosed with cancer between 1988 and 1994, the 5-year survival rate was 49%; among men in low poverty areas, the survival rate was 61%. Women in high and low poverty areas exhibited a similar gradient, with 5-year survival rates of 53 and 63%, respectively.

Differences also have been identified in cancer stage at diagnosis. In the same data set from SEER, both men and women in high poverty areas had a higher percentage of late-stage cancer diagnoses, for all four common cancer types. These gradients were

influenced by cancer site, most notably for prostate and female breast cancer (RR of distant-stage diagnosis was 1.9 and 1.7, respectively). Similar findings have been demonstrated in smaller population-based studies as well. For example, Roetzheim and colleagues studied all incident cases of colorectal, breast, and prostate cancer in Florida for 1 year. They found that uninsured and underinsured (i.e., Medicaid only) persons were significantly more likely to be diagnosed with late-stage disease *(36)*.

A side-note on data collection: Lantz and colleagues released a study in 2006 that explored the effect of race/ethnicity and "individual" socioeconomic factors on breast cancer stage at diagnosis *(37)*. As the authors pointed out, previous studies that examined the effect of SES on stage at diagnosis assessed SES at the geographic or contextual level, i.e., measuring SES based on census tract or zip code. The limitation of these types of studies is that broad-based inferences may not be appropriate substitutes for individual-level economic status. Changing the classification of SES may be useful, but this will require further evaluation. In the same way that expanding categories of race and ethnicity may potentially complicate data collection and analysis, expanding state and national cancer registries to report individual-level economic information may have a similar effect.

Finally, it is important to recognize that there are subgroups of this (socioeconomically disadvantaged) population that are frequently overlooked by most state and national cancer registries, as well as by most census data. These groups include the homeless, undocumented immigrants, and non-English speaking persons. These groups are typically on the fringes of health care, and they do not have access to insurance, preventative care, or cancer treatment. Not only are they more likely to be missed in population-based data collection but also they are less likely to be accessible, precluding them from benefiting from most prevention and screening programs. It is worthwhile to consider these groups, who are essentially disenfranchised in the health care system. As understanding of cancer epidemiology improves, and more effective preventive measures are introduced, it will be important to recognize and make provisions for these subgroups of the population.

The final population that will be addressed briefly is the elderly. There is a paucity of information in the literature about cancer epidemiology in the elderly, suggesting that older Americans are not featured prominently in these discussions. However, age-related disparities in screening, health care use, and access to care certainly exist, particularly among older racial/ethnic minorities, and the older poor.

Persons over the age of 65 now represent a significant portion of the population; as of the year 2000, an estimated one in every eight Americans is considered to be "elderly" *(38)*.

There is interesting overlap between the older adult population and other subpopulations. For example, in a pattern reflecting national change, the growth of the non-White elderly is expected to exceed that of elderly Whites in the next 50 years *(39)*. In addition, there is a disproportionate number of underinsured elderly in the United States. In 2002, the American Cancer Society estimated that 26% of the elderly have only Medicare coverage, compared with an estimated 16% of uninsured persons under age 65 *(40)*. Most of the disparities identified in this population are reflected in poorer health outcomes and inferior health care use and screening, as is discussed in **Subheading 5**. The rapid growth of this population has important implications for epidemiologic research. Furthermore, the substantial degree of overlap between the elderly, racial/ethnic minorities, and the economically disadvantaged can make it difficult to clarify the dynamics of existing disparities. The relative heterogeneity of this group challenges the scientific community to re-examine disparities among the elderly and possible ways to improve outcomes and access to care.

5. Epidemiology in Practice for Diverse Populations: Unique Issues of Investigative Measures and Population-based Interventions

An oversimplified summary of the goals of cancer epidemiology might be stated as such: First, the discipline seeks to better elucidate the distribution of the disease in certain populations, through research. Second, it strives to effectively implement preventive health care measures to control the disease. Best and colleagues expressed this dual function nicely: "A successful cancer control strategy must guide both the research endeavor and its application" *(41)*. Epidemiologic research is composed of two types of studies *(42)*. Descriptive studies identify broad patterns of cancer occurrence, and health-related factors as they pertain to cancer, in aggregate populations. By contrast, analytic studies actually measure associations between an exposure and the disease, by using data obtained from individuals. These "observational" studies make invaluable contributions to our understanding of cancer, and they serve as synergistic complements to experimental and clinical research.

Epidemiologic research has played an invaluable role in our public health—particularly over the latter half of the 20th century—offering insight into relationships between cancer and tobacco use, dietary factors, and even viral infections. However, certain subgroups of the population have historically been excluded—intentionally or unintentionally—from this research. For example, African Americans have been underrepresented in both occupational cancer *(43)* and cancer prevention *(44)*

research in the United States. Despite higher cancer-related mortality and lower survival rates, this minority group has historically not participated in preventive health studies *(45–47)*. This disparity is not limited to the African-American population. Data from the National Cancer Institute indicate that White Americans represent the overwhelming majority in preventive health trials, such as the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; accrual rates of other minority groups (including Asian and Native Americans, and those of Hispanic ethnicity) are two to three-fold lower than their respective cancer incidence *(9)*.

Several barriers to recruitment of minorities into research trials have been proposed, including mistrust of medical research by these populations *(48–50)*; lack of awareness of these trials *(48,51)*; lack of active recruitment by investigators *(52)*; language and cultural barriers *(53,54)*; and cost of participation *(48,55)*.

Mistrust of medical research exists among racial/ethnic minority groups, for a variety of reasons. For example, mistrust within the African American community has historical roots, dating back to the disclosure of the Tuskegee experiment. In addition, many racial/ethnic minorities embrace non-Western medical treatment, which may contribute to bidirectional mistrust between the patient and provider/investigator *(56)*. Last, a historical point of interest bears review. Until relatively recently in this country's history, the health status of African Americans (and, presumably, other racial/ethnic minorities) was not perceived to be a public health responsibility. Consider the essay *The Negro Health Problem in Southern Cities* released in 1915 by William Brunner, who was the health officer in Savannah, GA *(57)*. In this piece, in which Brunner addresses members of the American Public Health Association, he states: "It is up to the white people to … guard [African Americans] against tuberculosis, syphilis, etc … if [African Americans] are tainted with disease, [Whites] will suffer." Despite the condescending language, Brunner's argument for more inclusive public health efforts was actually quite forward-thinking. Although the idea of collective well-being seems obvious today, this was an innovative concept

Returning to possible explanations of low accrual of minorities into trials, lack of awareness of clinical research may be associated with lower education levels, particularly among communities with high levels of poverty *(59,60)*. Failure to actively recruit these populations may be related to reluctance on the part of investigators, or simple unawareness on the part of referring parties.

Language and cultural barriers are interesting factors in minority under-representation in epidemiologic research. For example, many indigenous and Native American languages do not have a word for "cancer"; this could conceivably make it difficult to effectively communicate the importance of the disease, and the relevance of preventative research. In addition, many U.S.

trials require proficiency of the English language, a prerequisite that would clearly preclude certain minorities from participating *(61,62)*.

A final factor that may play a role in the low accrual of poor minorities is the cost associated with trial participation. Although the costs associated with preventative trials may, theoretically, be less substantial than those associated with clinical treatment trials, costs related to provider visits and any required medical interventions may still apply. These expenses also could be a factor in the low accrual of certain subgroups, such as the uninsured. Of note, however, having insurance does not guarantee coverage of costs related to participation in a preventative clinical trial. According to the National Cancer Institute *(53,63)*, only 23 states currently have mandated provisions for coverage of clinical trial costs; of these states, not all insurers are held responsible (e.g., managed care plans only). Furthermore, not all states require coverage of preventative trials. Most coverage allowances are specifically for cancer *treatment* trials. It strikes an interesting chord that research emphasizing prevention is not given the same priority as that involving treatment.

After identifying the groups of the population affected by cancer disparity, recognizing the areas of disparity, and examining the lapses in recruiting these groups into preventative research, the final problem to address is disparity in access to cancer screening. Although there is no current widespread screening modality for lung cancer, there are available screening methods for other leading types of cancer, including colorectal and breast cancer. These tests include fecal occult blood testing, sigmoidoscopy/colonoscopy, and mammography. Many of the barriers identified in the previous discussion on research participation hold true for access to health care.

Differences in access to preventive health care have been identified in racial/ethnic minorities, the poor and underinsured, and the elderly. Specifically, inferior use of screening measures has been recognized in these groups. For example, African Americans—even those covered by Medicare—have been shown to be less likely to receive screening for colorectal cancer, including fecal occult blood testing and endoscopy *(64,65)*. Similar disparities have been identified in other racial/ethnic minorities. Native Americans face a distinctive disadvantage, in that the Indian Health Service—a major public health resource in that community—is underfunded by about 40% *(66)*. Non-white Hispanics are at increased risk of being uninsured; consequently, they are less likely to have a routine source of medical care *(67)*.

Other factors may contribute to disparate access to care among racial/ethnic minority groups, including linguistic disparities *(68)* and health-related risk behaviors. For example, certain racial/ethnic minority groups, including African

Americans and American Indians, have a higher prevalence of tobacco abuse, which is clearly a risk factor for many malignancies. In fact, tobacco products are disproportionately targeted to racial/ethnic minority communities *(69,70)*. In addition, one must consider the extent to which other high-risk health behaviors, such as alcohol abuse and poor diet, and traditional health beliefs play a role.

Cost and insurance issues clearly impact access to care and use of screening. Not surprisingly, subgroups of the U.S. population with low SES and low education level, such as the rural poor, are much less likely to receive preventative health measures, such as mammography and Pap tests *(71,72)*. An estimated 2.3 million eligible women of low SES did not receive recommended mammograms in 2003 *(73)*. Significant associations have been demonstrated between lack of insurance and decreased use of preventative health services, even among higher income groups *(74,75)*. Interestingly, even among insured individuals, unawareness of what screening options are covered by insurance might also be a barrier to care *(76)*. Cost-related obstacles may become even more important in the future, as screening and prevention advance to include genetic and biomarker testing *(77)*.

The elderly, in general, receive substandard medical care, and typically have poorer cancer-related mortality rates *(78)*. As noted, the most vulnerable elderly are those who are also racial/ethnic minorities, and/or poor or underinsured. Cancer screening rates are particularly low in these populations *(79–81)*. Issues of the relevance of screening in older populations invariably arise in scientific and policy-based discussions. However, consider these two points: First, if the elderly are *unable* to access care that is otherwise considered standard for the adult population, this suggests inequity in the public health system. Second, as the U.S. population ages and advances are made in cancer treatment, the applicability of these screening tests (and other preventative strategies) will likely need to be reconsidered.

6. Future Directions

The disparities addressed here have not gone ignored by the scientific community. In 1990, the Office of Research on Minority Health was established, under the supervision of Congress and the Director of the National Institutes of Health (NIH). Two years later, a $45 million health initiative was introduced that, in part, emphasized cancer research. Expanding upon this idea, the National Center on Minority Health and Health Disparities was established in 2000. The objectives of this center include pro-

motion of health disparities research; infrastructure development, and outreach to underserved populations. In 2003, the first *NIH Strategic Research Plan and Budget to Reduce and Ultimately Eliminate Health Disparities* was released. The Department of Health and Human Services has developed two additional initiatives, including the *Initiative to Eliminate Racial and Ethnic Disparities in Health* and *Healthy People 2010*. The latter proposal has set forth a goal of eliminating health disparities over the next decade, an ambitious objective to be sure.

It may not be realistic to expect that the disparities that have taken decades, even centuries, to develop, will be easily eradicated. However, initiatives such as those described in the paragraph above are necessary. In addition, legislation that supports institutions that serve these vulnerable populations, and that offers incentives to investigators committed to addressing health disparities, will continue to be crucial in the next several decades.

7. Conclusions

Regardless of the terminology that is used, or the definitions that are applied, disparities in the epidemiology of cancer are real. It is imperative that the scientific community recognizes these realities, and works to correct the flaws that enable them. These flaws are particularly prominent in the accrual of underserved populations into preventative research trials and the accessibility of preventative health measures. Addressing the lapses in the public health system will require a culturally-sensitive approach, and an interdisciplinary effort involving epidemiologists, clinical investigators, and public policy authorities.

References

1. Whaley AL. (2003) Ethnicity/race, ethics and epidemiology. *JNMA* **95(8),** 736–742.
2. Smedley, A (ed.) (1999) *R ace in North America: Origin and Evolution of a Worldview*, 2nd ed., Boulder, CO: Westview Press.
3. Smedley A, Smedley B. (2005) Race as biology is fiction, racism as a social problem is real: anthropological and historical perspectives on the social construction of race. *Am Psych* **60(1),** 16–26.
4. Williams DR. (1997) Race and health: basic questions, emerging directions. *Ann Epidemiol* **7(5),** 322–333.
5. Herman AA. (1996) Toward a conceptualization of race in epidemiologic research. *Ethn Dis* **6,** 7–20.
6. Bhopal R, Donaldson L. (1998) White, European, Western, Caucasian, or what? Inappropriate labeling in research on race, ethnicity, and health. *Am J Public Health* **88,** 1303–1307.
7. Fullilove MT. (1998) Comment: abandoning "race" as a variable in public health research—an idea whose time has come. *Am J Public Health* **88(9),** 1297–1298.
8. U.S. Census Bureau. (2000) CensusScope. http://www.censusscope.org/us/map_multiracial.html. Cited Jun 2007.

9. Hayes MA, Smedley BD (eds.) (1999) *The Unequal Burden of Cancer: An Assessment of NIH Research and Programs for Ethnic Minorities and the Medically Underserved.* Washington, DC: National Academies Press.
10. Oppenheimer GM. (2001) Paradigm lost: race, ethnicity, and the search for a new population taxonomy. *Am J Public Health* **91,** 1049–1055.
11. Frost F, Taylor V, Fries E. (1992) Racial misclassification of Native Americans in a surveillance, epidemiology and end results cancer registry. *J Nat Can Inst.* **84(12),** 957–62.
12. Trans-HHS Cancer Health Disparities Progress Review Group (2004) *Making Cancer Health Disparities History.* Submitted to the Secretary, U.S. Department of Health and Human Services, March 2004. Washington, DC: U.S. Department of Health and Human Services.
13. Thomson GE, Mitchell F, Williams M (eds.) (2006) Committee on the Review and Assessment of the NIH's Strategic Research Plan and Budget to Reduce and Ultimately Eliminate Health Disparities, *Examining the Health Disparities Research Plan of the National Institutes of Health: Unfinished Business.* Washington, DC: National Academies Press.
14. Krieger N. (2001) Theories for social epidemiology in the 21st century: an ecosocial perspective. *Int J Epidemiol* **30,** 668–677.
15. Vega J, Irwin A. (2004) Tackling health inequalities: new approaches in public policy. Bulletin of the World Health Organization, Geneva, Switzerland. http://www.who.int/bulletin. Cited Jun 2007.
16. Krieger N. (2005) Defining and investigating social disparities in cancer: critical issues. *Cancer Causes Control* **16(1),** 5–14.
17. Carter-Pokros O, Baquet C. (2002) What is a "health disparity"? *Public Health Rep* **117,** 426–434.
18. Whitehead M. (1991) The concepts and principles of equity and health. *Health Prom Int* **6(3),** 217–228.
19. U.S. Census Bureau. (2000) The Hispanic population in the United States: 1999. Current Population Reports, Series P20-527. Washington, DC: U.S. Census Bureau (http://www.census.gov/prod/2000pubs/p20-527.pdf). Cited Jul 2007.
20. Singh GK, Miller BA, Hankey BF, Edwards BK. (2003) *Area Socioeconomic Variations in U.S. Cancer Incidence, Mortality, Stage, Treatment, and Survival, 1975–1999.* National Cancer Institute, Cancer Surveillance Monograph Series, Number 4. National Institutes of Health Publication No. 03–5417. Bethesda, MD: National Cancer Institute.
21. Miller B, Kolonel L, Bernstein L, et al. (eds.) (1996) *Racial/Ethnic Patterns of Cancer in the United States 1988–1992.* National Institutes of Health Publication No. 96–4104. Bethesda, MD: National Cancer Institute.
22. Wingo PA, Ries LA, Rosenberg HM, et al. (1998) Cancer incidence and mortality, 1973–1995: a report card for the U.S. *Cancer* **82,** 1197–1207.
23. Howe HL, Wu X, Ries LA. (2006) Annual report to the nation on the status of cancer, 1975-2003, featuring cancer among U.S. Hispanic/Latino populations. American Cancer Society. Published online September 2006 in Wiley InterScience (www.interscience.wiley.com). Cited Jul 2007.
24. Piffath TA, Whiteman MK, Flaws JA, et al. (2001) Ethnic differences in cancer mortality trends in the U.S., 1950–1992. *Ethn Health* **6(2),** 105–119.
25. Landis SH, Murray T, Bolden S, et al. (2000) Cancer statistics, 2000. *CA Cancer J Clin* **50(1),** 2398–2424.
26. Clegg LX, Li FP, et al. (2002) Cancer survival among US Whites and minorities. *Arch Int Med* **162,** 1985–1993.
27. Haddock RL, Naval CL. (1996) Cancer in Guam: a review of death certificates from 1971–1995. *Pac Health Dial* **4(1),** 66–75.
28. Nasca PC, Pastides H (eds.) (2001) *Fundamentals of Cancer Epidemiology.* Sudbury, MA: Jones and Bartlett Publishers.
29. Ni H, Cohen R. (2002) Trends in health insurance coverage by race/ethnicity among persons under 65 years of age: United States, 1997–2001. National Center for Health Statistics Health E-Stats. http://www.cdc.gov/nchs/products/pubs/pubd/hestats/healthinsur.htm. Cited Jul 2007.
30. Grann VR. (2003) Health insurance and cancer survival. *Arch Int Med.* **163,** 2123–2124.
31. Schwartz KL, Crossley-May H, Vigneau FD, et al. (2003) Race, socioeconomic status and stage at diagnosis for five common malignancies. *Cancer Causes Control* 14, 761–766.
32. Watlington AT, Byers T, Mouchawar J, et al. (2007) Does having insurance affect differences in clinical presentation between Hispanic and non-Hispanic white women with breast cancer? *Cancer* **109(10),** 2093–2099.
33. Du XL, Meyer TE, Franzini L. (2007) Meta-analysis of racial disparities in survival in association with socioeconomic status

among men and women with colon cancer. *Cancer* **109,** 2161–2170.
34. Doubeni CA, Field RS, Buist DS, et al. (2007) Racial differences in tumor stage and survival for colorectal cancer in an insured population. *Cancer* **109,** 612–620.
35. Williams DR, Jackson PB. (2005) Social sources of racial disparities in health. *Health Aff* **24(2),** 325–334.
36. Roetzheim RG, Pal N, Tennant C, et al. (1999) Effects of health insurance and race on early detection of cancer. *J Nat Canc Inst.* **91(16),** 1409–1415.
37. Lantz PM, Mujahid M, Schwartz K, et al. (2006) The influence of race, ethnicity, and individual socioeconomic factors on breast cancer stage at diagnosis. *Am J Public Health* **96(12),** 2173–2178.
38. Hetzel L, Smith A (eds.) (2001) *The 65 Years and Over Population: 2000. Report No. C2KBR/01–10.* Economic and Statistic Administration, Bureau of the Census. Washington, DC: U.S. Department of Commerce.
39. U.S. Census Bureau. (1993) We the American Elderly. Washington, DC: U.S. Department of Commerce.
40. American Cancer Society. (2002) Cancer facts and figures. Washington, DC: American Cancer Society.
41. Best A, Hiatt RA, Cameron R, et al. (2003) The evolution of cancer control research: an international perspective from Canada and the United States. *Cancer Epidemiol Biomark Prev* **12**, 705–712.
42. Thun EE, Jemal AM. (2006) Cancer epidemiology. In: *Cancer Medicine*, 7th ed. Kufe DW, Bast RC, Hait WN, et al. (eds.) London, UK: BC Decker, pp. 339–353.
43. Zahm SH, Pottern LM, Lewis DR, et al. (1994) Inclusion of women and minorities in occupational cancer epidemiology research. *J Occ Med* **36,** 842–847.
44. Henderson C. (1997) Meeting planned to encourage minority involvement in trials. *Can Wkly Plus* **2,** 10.
45. Shavers VL, Lynch CF, Burmeister LF. (2002) Racial differences in factors that influence the willingness to participate in medical research studies. *Ann Epidemiol* **12,** 248–256.
46. Goldberg KB, Goldberg P (eds.) (1997) Minority accrual low in prevention trials: NCI plans programs to boost enrollment. *Cancer Lett* **22,** 1–4.
47. Coleman EA, Tyll L, LaCroix AZ, et al. (1997) Recruiting African American older adults for a community health promotion intervention: which strategies are effective? *Am J Prev Med* **13(Suppl 6),** 51–56.
48. Harris Y, Gorelick PB, Samuals P, et al. (1996) Why African Americans may not be participating in clinical trials. *J Nat Med Assoc* **88,** 630–634.
49. Shavers-Hornaday VL, Lynch CF, Burmeister LF, et al. (1997) Why are African Americans under-represented in medical research studies? Impediments to participation. *Ethn Health* **2,** 31–46.
50. Brandon, D.T., L.A. Isaac, and T.A. LaVeist. (2005) The legacy of Tuskegee and trust in medical care: is Tuskegee responsible for race differences in mistrust of medical care? *J Natl Med Assoc* **97(7),** 951–956.
51. Comis, R., et al. (2000) A Quantitative Survey of Public Attitudes Towards Cancer Clinical Trials. http://www.cancersummit.org/media_news/harris%20poll.pdf. Cited Mon 2008
52. Wendler, D., et al. (2006) Are racial and ethnic minorities less willing to participate in health research? *Public Lib Sci Med* **3(2),** e19.
53. National Cancer Institute (2002). States That Require Health Plans to Cover Patient Care Costs in Clinical Trials. http://www.cancer.gov/clinicaltrials/developments/laws-about-clinical-trial-costs. Cited Jun 2007.
54. Intercultural Cancer Council: Cancer Fact Sheets (2004) http://iccnetwork.org/cancerfacts/. Cited Jul 2007.
55. Baquet CR, Horn JW, Gibbs T, et al. (1991) Socioeconomic factors and cancer incidence among blacks and whites. *J Natl Can Inst* **93,** 551–557.
56. Frank-Stromberg M, Olsen SJ (eds.) (2001) *Cancer Prevention in Diverse Populations: Cultural Implications for the Multidisciplinary Team*, 2nd ed. Pittsburgh, PA: Oncology Nursing Society.
57. Brunner WF. (1915) The Negro health problem in southern cities. *Am J Public Health* **5(3)**, 183–190.
58. Shavers VL, Shavers BS. (2006) Racism and health inequity among Americans. *JNMA* **98(3),** 386–393.
59. Burhansstipanov L. (1996) Overcoming psycho-social barriers to Native American cancer screening research. In: *Conferences Summary: Recruitment and Retention of Minority Participants in Clinical Cancer Research.* National Cancer Advisory Board, National Cancer Institute. U.S. Department of Health and Human Services. NIH 96-4182, pp. 109–127.
60. Borrayo EA, Lawsin C, Coit C. (2005) Latinas' appraisal of participation in breast cancer prevention clinical trials. *Cancer Control* **12(Suppl 2),** 107–110.

61. Nguyen TT, Somkin CP, Ma Y. (2005) Participation of Asian-American women in cancer chemoprevention research: physician perspectives. *Cancer* **104(12 Suppl),** 3006–3014.
62. Giuliano AR, Mokuau N, Hughes C, et al. (2000) Participation of minorities in cancer research: the influence of structural, cultural, and linguistic factors. *Ann Epidemiol* **10(8) Suppl 1,** S22–S34.
63. National Cancer Institute Cancer Clinical Trials: The Basic Workbook. (2002) NIH Publication no. 02-5050. http://www.cancer.gov/PDF/091e02f3-5bb9-4ba6-a988-9ad35eab6a61/BasicsWorkbook_m.pdf. Bethesda, MD: National Institutes of Health. Cited Jul 2007.
64. Schenck AP, Klabunde CN, Davis WW. (2006) Racial differences in colorectal cancer test use by Medicare consumers. *Am J Prev Med* **30(4),** 320–326.
65. O'Malley AS, Forrest CB, Feng S, et al. (2005) Disparities despite coverage: gaps in colorectal cancer screening among Medicare beneficiaries. *Arch Int Med.* **165(18),** 2129–2135.
66. Burhansstipanov L, Olsen S. (2001) Cancer prevention and early detection in American Indian and Alaska Native populations. In: *Cancer Prevention in Diverse Populations: Cultural Implications for the Multidisciplinary Team*, Frank-Stromborg M, Olsen SJ (eds.). St. Louis, MO: Mosby Publishers.
67. American College of Physicians–American Society of Internal Medicine. (2000) No health insurance? It's enough to make you sick. Latino community at great risk. Philadelphia: College of Physicians–American Society of Internal Medicine; White Paper. Philadelphia, PA: American College of Physicians–American Society of Internal Medicine.
68. Ponce NA, Hays RD, Cunningham WE. (2006) Linguistic disparities in health care access and health status among older adults. *J Gen Int Med.* **21(7),** 786–791.
69. Centers for Disease Control and Prevention. (2000) National Center for Chronic Disease Prevention and Health Promotion. Tobacco information and prevention source (Tips)-Hispanics and tobacco. http://www.cdc.gov/tobacco/data_statistics/sgr/sgr_1998/sgr-min-hsp.htm. Cited July 2007.
70. U.S. Department of Health and Human Services. (1998) Tobacco use among US racial/ethnic minority groups—African Americans, American Indians and Alaska Natives, Asian Americans and Pacific Islanders, and Hispanics: a report of the Surgeon General. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.
71. Friedell G, Brown P. (2000) Report of the data working group meeting–issues for rural, poor White, medically underserved (Appalachian) populations. Intercultural Cancer Council Report, 26 Sep 2000.
72. Fernandez ME, DeBor M, Candreia MJ, et al. (1999) Evaluation of ENCOREplus. A community-based breast and cervical cancer screening program. *Am J Prev Med* **16(3 Suppl),** 35–49.
73. Tangka FK, Dalaker J, Chattopadhyay SK, et al. (2006) Meeting the mammography screening needs of underserved women: the performance of the National Breast and Cervical Cancer Early Detection Program in 2002–2003 (United States). *Cancer Causes Control* **17(9),** 1145–1154.
74. Ross JS, Bradley EH, Busch SH. (2006) Use of health care services by lower-income and higher-income uninsured adults. *JAMA* **295(17),** 2027–2036.
75. Rosenberg L, Wise LA, Palmer JR, et al. (2005) A multilevel study of socioeconomic predictors of regular mammography use among African-American women. *Cancer Epidemiol Biomark Prev* **14(11 Pt 1),** 2628–2633.
76. McAlearney AS, Reeves KW, Tatum C, et al. (2005) Perceptions of insurance coverage for screening mammography among women in need of screening. *Cancer* **103,** 2473–2480.
77. Giacomini M, Miller F, O'Brien BJ. (2003) Economic considerations for health insurance coverage of emerging genetic tests. *Community Genet* **6(2),** 61–73.
78. Hodgson, DC, Fuchs, CS, Ayanian, JZ. (2001) Impact of patient and provider characteristics on the treatment and outcomes of colorectal cancer. *J Natl Canc Inst* **93(7),** 501–515.
79. Centers for Disease Control and Prevention. (2003) The National Breast and Cervical Cancer Early Detection Program: Reducing Mortality through Screening. Atlanta, GA: National Center for Chronic Disease Prevention and Health Promotion.
80. Mandelblatt J, Gold K, O'Malley A, et al. (1999) Breast and cervix cancer screening among multiethnic women: role of age, health, and source of care. *Prev Med* **28,** 418–425.
81. Federal Interagency Forum on Aging-Related Statistics. (2000) Older Americans 2000: Key Indicators of Well-Being. Hyattsville, MD: U.S. Government Printing Office.

Chapter 5

Epidemiology of Multiple Primary Cancers

Isabelle Soerjomataram and Jan Willem Coebergh

Summary

Cancer patients have a 20% higher risk of new primary cancer compared with the general population. Approximately one third of cancer survivors aged >60 years were diagnosed more than once with another cancer. As the number of cancer survivors and of older people increases, occurrence of multiple primary cancers is also likely to increase. An increasing interest from epidemiologic and clinical perspectives seems logical. This chapter begins with the risk pattern of multiple cancers in the population of a developed country with high survival rates. Multiple cancers comprise two or more primary cancers occurring in an individual that originate in a primary site or tissue and that are neither an extension, nor a recurrence or metastasis. Studies of multiple cancers have been mainly conducted in population-based settings, and more recently in clinical trials and case control studies leading to further understanding of risk factors for the development of multiple primary cancers. These factors include an inherited predisposition to cancer; the usual carcinogenic or cancer-promoting aspects of lifestyle, hormonal, and environmental factors; treatment of the previous primary cancer; and increased surveillance of cancer survivors. Finally, implication on research strategies and clinical practice are discussed, covering the whole range of epidemiologic approach.

Key words: Lifestyle factors, multiple cancers, risk factors, survival.

1. Introduction

The number of cancer survivors has been increasing, with the rising survival tripling between 1971 and 2002 in the United States. Individuals who were affected by cancer have a higher risk of subsequent primary cancers either in the same organ or in another organ. Therefore, the prevalence of patients with multiple cancers is expected to continue increasing. Accordingly, in the United States, an increased rate of new cancer diagnosis was observed among cancer survivors diagnosed in the most recent period (relative risk [RR] =1.21 for those diagnosed from 1995 to 2000 vs. 1.14 for those diagnosed

Mukesh Verma (ed.), *Methods in Molecular Biology, Cancer Epidemiology, vol. 471*
© 2009 Humana Press, a part of Springer Science + Business Media, Totowa, NJ
Book doi: 10.1007/978-1-59745-416-2

from 1990 to 1994) *(1)*. Likewise, the overall risk of subsequent malignancies among cancer survivors in Finland increased 50% from the 1950s to the 1980s *(2)*.

Overall, about 8% of newly diagnosed cancers are in individuals who already have had a previous primary cancer *(3)*. Thus, annually we expect almost 900,000 new multiple cancer cases (8% of 10.9 million *(4)* new cases) worldwide. This number tends to increase among others because the growing proportion of the elderly whose prevalence of multiple cancer is highest: only 5–12% of cancer patients aged 50–64 years were previously diagnosed with cancer, versus 12–26% among those aged >80 years *(5)*. Additionally, other forces increase the frequency of multiple cancers such as awareness of such cancers and use and sensitivity of screening and better treatment. Conversely, the diagnosis of a new primary cancer may confer to higher mortality *(6)* and reduce quality of life of cancer survivors. Therefore, the phenomenon of multiple cancers has become of increasing epidemiologic and clinical interest.

2. Definitions

Multiple cancers are defined as two or more primary cancers occurring in an individual that originate in a primary site or tissue and that are neither an extension, nor a recurrence, nor metastasis (International Agency for Research on Cancer). They may occur in the same tissue or organ or affect different tissues or organs. Multiple cancers can be categorized into 1) synchronous, in which the cancers occur at the same time (no common rule exists, it may be 2, 3, or 6 months or even 1 or 2 years); and 2) metachronous, the subsequent cancer occurs after the period covered for synchronous cancer *(7)*. Most cancer registries adopt either the criteria suggested by the International Agency for Research on Cancer or of the Surveillance, Epidemiology, and End Results Program, the former being more strict, resulting in fewer multiple cancer cases *(7–9)*. The term second (primary) cancer will be used often in the discussion of multiple malignancies and also in this chapter, because currently 75% of all multiple cancer cases were cancer survivors diagnosed with a second primary cancer *(3)*. To distinguish a new primary tumor from recurrent tumors or metastatic lesions can be sometimes problematic, and they may lead to misclassification, especially in paired organs such as the breast or in organs having the same morphology, such head and neck cancers (squamous cell carcinoma) or urinary tract cancers *(10,11)* (urothelial cell carcinoma). Distinguishing these entities is not only important to accurately assess the risk of multiple

cancers but also to determine appropriate treatment *(7, 12)*. Tumour characteristics, i.e. histology, location, stage, or molecular signature provide guidelines in their categorization *(12–14)*.

3. Epidemiology

Nowadays, about one in every six cancer survivors had once breast cancer. Thus, women with breast cancer also represent the largest proportion (29%) of the total multiple cancer prevalence reported in the United States in 2002 (**Fig. 5.1**; *(3)*). The second and third largest group of multiple cancer cases include men and women whose first primary cancer was colorectal cancer (17%) and men whose first primary cancer was prostate cancer (14% *(3)*). Generally, women have a slightly higher RR of a second cancer compared with men (17% higher for women vs.11% for men), probably because common female cancers confer much better survival chances compared with that of males *(1)*.

The percentage of multiple cancer cases is highest among the elderly, peaking among those aged 70–79 years, i.e., 36 vs. 4, 9, 19, and 30% at 0–49, 50–59, 60–69, and 80^{+} years, respectively (**Fig. 5.2a**; *(3)*). However, when risk of a second cancer is compared with cancer risk in the population, a striking trend was observed from the youngest to the oldest age group. This ranges from a RR of 6 among those first diagnosed with a cancer at the age of 0–17 years decreasing to 0.92 among those diagnosed with first cancer older than 80 years (**Fig. 5.2b**; *(1)*). Underreporting of second cancers and a shorter life expectancy in the highest age group may cause the observed lower absolute and RR.

Studies of geographic difference in incidence of multiple primaries have been hampered by the different registration methods used by various cancer registries *(15, 16)*. In an international

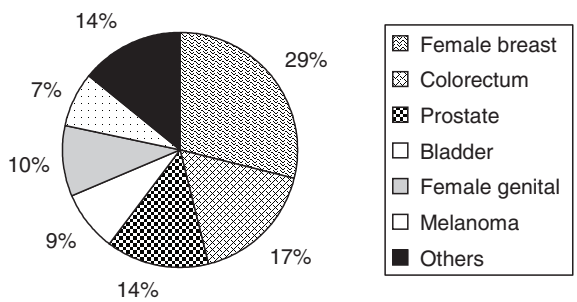


Fig. 5.1. Proportion newly diagnosed multiple cancer cases according to first primaries.

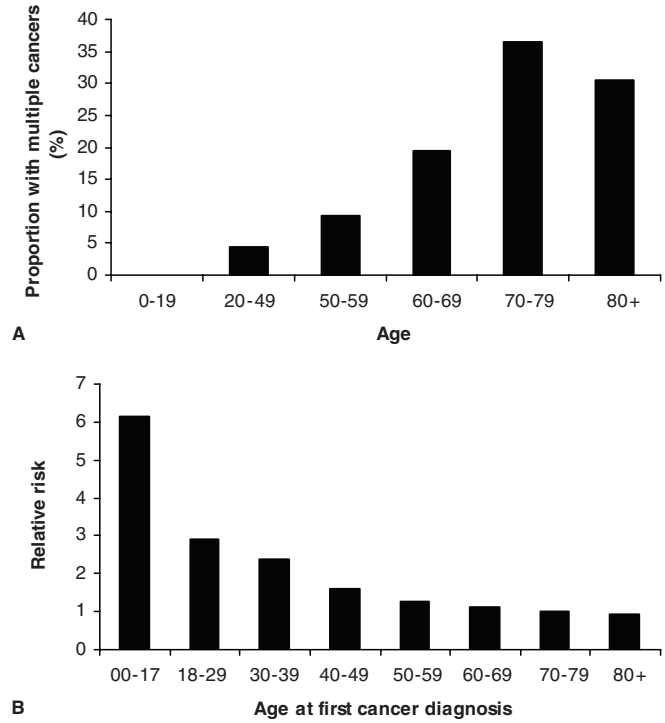


Fig. 5.2. Proportion of cancer patients with multiple cancers and relative risk of subsequent primary cancers according to age.

context with common registration methods multiple cancer studies present a unique opportunity to assess the etiology of cancer *(15, 17–20)*. Using the data of 13 worldwide cancer registries, the RR of second cancer was compared in patients with previous skin cancers in sunny countries compared with less sunny countries *(20)*. Skin cancers increased with higher sun exposure, whereas the UV of sunlight through the production of vitamin D might protect against some cancers *(21, 22)*. Concordantly, this study reported a significantly lower RR of all second solid primary cancers (except skin and lip) after skin cancer in the sunny countries compared to the less sunny countries.

4. Study Design and Method of Analysis of Multiple Cancer Studies

4.1. Cohort Study

In a cohort study, also referred as longitudinal study, a large group of cancer patients is followed forward in time to ascertain the occurrence of another primary cancer. It can be done in two ways: 1) prospective cohort study, in which the cohort is identified in the present and followed into the future; or 2) historical cohort

study (often denoted as retrospective cohort study), in which the cohort is identified in the present (some may already exhibit the outcome of interest, in this context another primary cancers) and by means of medical records the cancer experience is reconstructed between the defined time in the past and the present *(23–25)*. Data sources include population-based cancer registries, hospital-based cancer registries, and clinical trial series. Population-based study gives large number of cases, allowing detection of even a small increase in risk as well of having determined reference population. However, detailed data on risk factors are usually lacking. Hospital-based studies may comprise large and extensive data, clinical trial databases even in more detail, but numbers of cases are usually smaller and occurrence of rare second cancer becomes much harder to ascertain.

To determine whether cancer patients are at a higher or lower risk of developing cancers than the general population, the incidence of subsequent cancers among these patients (observed incidence) is compared with the incidence of such cancers in the general population. The expected incidence derived from calculating person-years of follow-up in the cohort stratified by gender, age, and calendar year. Dividing the observed incidence by the expected incidence results in the standardized incidence ratio (SIR) *(23,24,26,27)*. Examining SIR and its significance is a way to exclude the role of chance in assessing the risk of second primary cancer. Categorizing SIR in different follow-up times (after diagnosis of first cancers) may give clues to the excess cases due to heightened medical surveillance or to the role of cancer treatment. Excess risk only during the first years after first cancer diagnosis suggests a surveillance bias. Excess risk of solid second tumor occurs only after a latency time of 5–10 years. Whether a subsequent cancer has a large burden in a cohort, absolute excess risk provides the measure of the excess number of subsequent malignancies per 10,000 patients per year *(23,25)*. It is estimated by subtracting the expected number of second cancers from the observed number, and dividing this by the number of person-years, usually per 10,000 cases. The last confers the cumulative risk, which is the proportion of patients who would develop a subsequent cancer conditional on survival. Cumulative risk can be calculated either using actuarial method *(28)* or cumulative incidence function *(29)*.

When time trend of multiple cancers is the main study interest, one should adjust for factors that increase the risk of subsequent cancers such as length of follow-up after the diagnosis of the first primary. A fixed inception cohort method where risk of second cancer in different cohorts with the same follow-time is compared may overcome this problem *(30)*. Multivariate regression adjusting for various factors may be used to study determinants of interest corrected for confounding factors *(31)*.

4.2. Case-Control Study

A nested case-control study within a cohort presents the opportunity to assess the role of risk factors in greater detail, such as cancer therapy *(23,24,32)* or behavioral risk factors *(33)*. Detailed data could be ascertained through medical records or questionnaire to the therapist or patients themselves. Cases are patients with second cancer (or more). Controls are patients who do not develop the second cancer, randomly matched by age, gender, calendar year of diagnosis, and naturally, length of follow-up time. Relative risk can then be calculated by comparing different exposures of interest among the cases and the cohort. Overmatching (of nonconfounding factor[s]) would unnecessarily reduce statistical power and finally produce nonassociation of exposure and cases *(24)*.

5. Causes

Multiple cancers arise in the same individual due to several following causes: 1) host factors such as genetic or hormonal factors, 2) lifestyle, 3) first cancer treatment, and 4) environment. In most patients, a combination of several factors likely contributes to the occurrence of multiple cancers (**Fig. 5.3**; *(32,34)*). Additionally, an elevated risk of multiple malignancies may also be caused by higher medical surveillance after a cancer diagnosis or merely due to chance (see study method to assess the role of risk factors).

5.1. Genetic Predisposition

About 5–10% of all cancers arise in individuals with an inherited genetic mutation conferring to heightened cancer-specific susceptibility *(35)*. A short list of selected cancer inherited syndromes, their gene mutations, and penetrance based on comprehensive review in these area are listed in **Table 5.1** *(35–37)*. Increased risk of multiple cancers has been consistently reported among patients with family history of cancer *(38–44)*. Cancer in patients

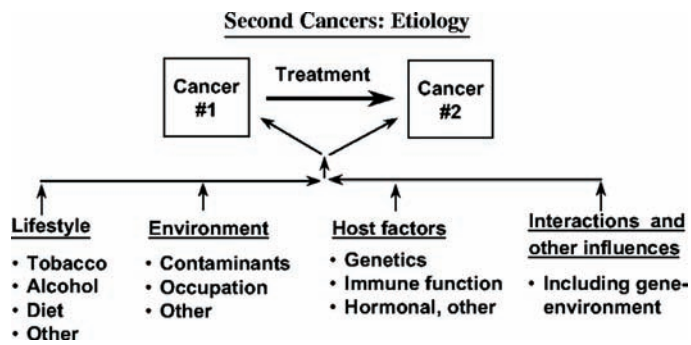


Fig. 5.3. Factors related with risk of subsequent primary cancers.

Table 5.1
Selected inherited cancer syndromes, reported in various multiple cancer cases (modified table from Fearon et al. (*36*) and Nagy et al. (*37*))

Syndrome	Affected sites	Penetrance (%)	Gene(s)
Familial breast cancer	**Breast**, ovary, male breast, pancreas, prostate, melanoma	Up to 85	BRCA1, BRCA2
HNPCC or Lynch syndrome	**Colorectum**, corpus uteri, ovary, hepatobiliary and unrinary tract, brain. Also Muir-Torre and Turcot variant-related tumors.	90	MLH1, MSH1, MSH2, PMS1, PMS2
Hereditary retinoblastoma	**Eyes**, bone and soft tissue sarcoma	90	RB
Li-Fraumeni syndrome	**Sarcoma, breast**, brain, leukaemia and adrenocortical cancer	90–95	TP53
Cowden syndrome	**Breast**, thyroid corpus uteri	~50	PTEN (MMAC1)
Familial melanoma	**Melanoma**, pancreas	~90	CDKN2A (p16)
Multiple endocrine neoplasia type 1	**Parathyroid**, entero-pancreas, pituitary	95	MEN1
FAP	**Colorectum**, thyroid, pancreas, liver, central nervous system, and other benign conditions	~100	APC

Bold refers to most affected sites (highest penetrance). More detailed table and overview please refer to **refs.** *35–37* and *50*.
Adapted from Nagy, R., Sweet, K., Eng, C. (2004) Highly penetrant hereditary cancer syndromes. Oncogene **23,** 6445–70 with permission from Nature Publishing Group and from Fearon, E. R. (1997) Human cancer syndromes: clues to the origin and nature of cancer. Science **278,** 1043–50 with permission from American Association for the Advancement of Science.

with heritable cancer susceptibility generally presents at early ages *(42,44)*. Among female breast cancer patients aged younger than 50 years, risk of ovarian cancer is four-fold compared with those with breast cancer diagnoses older than 50 years *(19,45–47)*. This is consistent with the presence of germline mutation BRCA1/2 and possibly also other mutations. In 10 years after breast cancer diagnosis, 20% *(48)* to 30% *(49)* of patients carrying this mutation would be diagnosed with ovarian or contralateral breast cancer, respectively. Breast cancer patients younger than 45 years also may carry a germline mutation in the TP53 (Li-Fraumeni syndrome), which also is related to higher incidence of soft tissue and sarcoma as well as brain tumors, adrenal cortical carcinoma, and leukemia.

Early onset of colon cancer has been associated to Lynch syndrome (hereditary nonpolyposis colorectal cancer [HNPCC]) or familial adenomatous polyposis (FAP). In addition, HNPCC patients have a heightened risk of endometrial, ovarian, stomach, and kidney cancer *(35,50)*. Moreover, colorectal cancer patients also may have a higher breast cancer risk through inherited mutations of CHEK2 *(51)*.

Beside the role of single-gene mutation, the polygenic model may explain part of the increased risk of multiple cancers in an individual *(8,52,53)* also interacting with other risk-enhancing factors, such as smoking and alcohol, or cancer treatment, such as radiation. A polygenic model would explain the occurrence of cases with familial clustering of cancers without (detectable) specific germline mutation, i.e., only 70% of all families meeting the criteria of Li-Fraumeni syndrome showed a germline mutation in TP53 *(35,42)*.

5.2. Common External Factors

5.2.1. Smoking and Alcohol

Smoking- and alcohol-related second malignancies account for 35% of the total excess of subsequent cancer cases observed in the U.S. cancer survivors *(1)*. Most multiple cancer studies have used population-based registry, and data on behavioral risk factors such as smoking or alcohol intake are not readily available. The clustering of smoking- or alcohol-related cancer in an individual designates them as common risk factors. The pattern of multiple cancers risk that share common etiologic factors is a useful tool to give insight of their etiology, and it should generally comply to the following rules: 1) significant reciprocal increased risk (increased risk of cancer A after cancer B and vice versa); 2) persistent increased of relative risk since diagnosis of the index tumor; and 3) role of first cancer treatment could be excluded, i.e., similar risk pattern between those who received surgical and radiotherapy *(27)*. A consistent excess of subsequent primary smoking-related cancers (i.e., oral cavity, pharynx, pancreas, larynx, lung, kidney, and bladder) has been reported among patients ever diagnosed with similar cancers *(18,54–57)*. And for alcohol, it is likely to have contributed to the increased risk of liver and esophageal cancer among laryngeal cancer patients *(57,58)*.

Where individual behavioral history is available, cigarette smoking clearly increased risk of smoking-related second primary cancer (**Table 5.2**; *(58–63)*). Patients with laryngeal/hypopharyngeal carcinoma who smoked exhibited a hazard ratio of 3 to 4 for lung cancer compared with those who never smoked or smoked only occasionally. Furthermore, studies indicated that smoking cessation after cancer diagnosis lowers the risk of new smoking-related malignancies *(60,64)*. Likewise, laryngeal and hypopharyngeal cancer survivors with the highest alcohol consumption (≥121 g/day) exhibited a three-fold higher risk of upper aerodigestive tract cancers compared with those with the lowest alcohol intake

Table 5.2
Risk of lung and upper aerodigestive tract cancer after laryngeal/hypopharyngeal carcinoma, according to smoking and alcohol intake index *(58)*

Site of second primary/risk factor	Lung			UADT[a]			
	Cases/total cohort	HR	95% CI		Cases/total cohort	HR	95% CI
Pack-yr of cigarette smoking				**Avg alcohol drinking (g/day)**			
0–20	3/115	1		0–40	4/197	1	
21–40	26/362	3.3	0.9–11.0	41–80	4/227	0.8	0.2–3.3
41–60	15/264	2.4	0.7–8.6	81–120	12/206	3.0	0.9–9.5
≥61	10/128	3.9	1.0–14.6	≥121	17/246	3.5	1.1–11.2
Stratified log-rank test, *P* value			0.06				0.003

[a]UADT, upper aerodigestive tract, including lip, tongue, oral cavity, oropharynx, nasopharynx, hypopharynx, larynx, esophagus (ICD-9 140-150).
Data was based on 876 male primary larynx and hypopharynx cancer patients. HR was adjusted for age, occupational group, alcohol drinking, cigarette smoking, and site of first cancer (hypopharynx or larynx) (*see* **ref**. *58)*. Copyright @2005 American Cancer Society. This material is reproduced with permission of Wiley-Liss, Inc., a subsidiary of John Wiley & Sons, Inc.

(0–40 g/day) (**Table 5.2**; *(58)*). Breast cancer patients with the highest alcohol intake exhibited almost a two-fold higher risk of colorectal cancer compared with nondrinkers *(40)*.

5.2.2. Diet and Obesity

Dietary factors, including obesity, diet low of fruit and vegetable, and diet high in fat, accompanied by low physical activity have been related to occurrence of a large number of cancers in the general population *(65,66)*, and they also seem to account for multiple cancers involving the breast, female reproductive organs, and lower and upper digestive tract *(19,40,45–47,67)*. Obese breast cancer patients had approximately 2-fold greater hazard of contralateral breast tumors relative to underweight/normal-weight women (**Table 5.3**; *(40,68,69)*). Likewise, obesity, adult weight gain, or both increased the risk of other second primary cancers, including endometrial and colon cancer risk (**Table 5.3**; *(40,41,69)*). Furthermore, very obese colon cancer patients (body mass index [BMI] ≥ 35 kg/m^2) showed greater risk of a recurrence or second primary tumor of the colon; hazard ratio [HR] = 1.38, 95% confidence interval [CI] = 1.10–1.73) than normal-weight patients (BMI = 18.5–24.9 kg/m^2) *(70)*. Finally, high-citrus fruit and vegetable consumption may reduce lung cancer risk among patients previously diagnosed with laryngeal cancer by 10–60% *(58)*.

Table 5.3
Risk of primary breast, endometrial, and colorectal cancer after breast cancer, according to BMI before or at the diagnosis of first primary breast cancer

Site of second primary/risk factor	Breast[a]		Endometrial[b]			Colorectal[b]	
	HR	95% CI		HR	95% CI	HR	95% CI
BMI (kg/m^2)			BMI[c] (kg/m^2)				
≤24.9	1		< 22.5	1		1	
25.0–29.9	1.22	0.87–1.71	22.5–25.0	0.98	0.50–1.90	0.91	0.51–1.60
≥30.0	1.58	1.10–2.25	25.1–28.8	1.07	0.55–2.07	1.54	0.92–2.59
			≥ 28.9	2.23	1.23–4.05	1.67	0.99–2 .82

[a]Data was based on 193 newly diagnosed patients with contralateral breast cancer among 3,385 primary (early stage) breast cancer patients. Hazard ratio was adjusted for treatment, age, menopausal status, race, tumor size, estrogen receptor level, and progesterone receptor level. Risk of contralateral breast cancer was similarly elevated by increasing BMI in women who were premenopausal and postmenopausal at study entry *(68)*. Adapted from Dignam, J. J., Wieand, K., Johnson, K. A., Fisher, B., Xu, L., Mamounas, E. P. (2003) Obesity, tamoxifen use, and outcomes in women with estrogen receptor-positive early-stage breast cancer. J Natl Cancer Inst. **95,** 1467–76. With permission from Oxford University Press.

[b]Data was based on 90 primary endometrial cancer cases among 5,724 postmenopausal breast cancer patients and 127 primary colorectal cancer cases among 8,020 postmenopausal breast cancer patients. Regression models conditional on age and hazard ratio was adjusted for year of diagnosis, stage of breast cancer at initial diagnosis, family history of breast cancer, pack-years of cigarette smoking, recent alcohol intake, parity, and postmenopausal hormone therapy *(40)*.

[c]BMI is weight in kilograms divided by height in meters squared.Adapted from Trentham-Dietz, A., Newcomb, P. A., Nichols, H. B., Hampton, J. M. (2006) Breast cancer risk factors and second primary malignancies among women with breast cancer. Breast Cancer Res Treat. With permission from Springer Netherlands.

5.2.3. Hormonal Factor

Hormonal and reproductive factors (age at birth of first child, at menarche and menopause and parity; use of oral contraception and hormone-replacement therapy (HRT)) are related to risk of several cancers, such as breast, endometrial, ovarian, and colon cancers. Cancer patients who had premature menopause, i.e., due to chemotherapy, exhibited a lower risk of second primary breast cancer *(71)*. Breast cancer survivors who were younger at menarche and had fewer children showed an increased risk of second primary breast cancer *(33,40)*. Furthermore a reduced colorectal cancer risk is observed among those who were older at menarche, younger at menopause, and used HRT *(40)*. The role of reproductive factors did not seem to alter the effect of first cancer treatment *(38,71,72)*.

5.2.4. Infection and Immunosuppression

Several infectious agents are considered to be causes of cancer in humans. Human papillomavirus (HPV) infection is likely to play a role in case of multiple cancers of the tonsil, oropharynx,

esophagus, anus, cervix uteri, vagina, and vulva. Among patients with cancer of the cervix, vulva, and vagina, a three- to five-fold increased risk of other HPV-related cancers was observed *(73–75)*. A similar excess of HPV-related cancers was reported in women diagnosed with in situ cervical cancer *(74)*. Analyzing the risk for in situ and invasive cervical cancer patients separately gives the opportunity to assess and exclude the role of cancer treatment i.e. radiation is given only to patients with invasive cervical cancer. Furthermore, identical HPV DNA integration loci in tissue from the initial cervical cancer and in the subsequent vaginal or vulvar cancers have been detected, indicating common etiology *(76)*.

Epstein–Barr virus (EBV) has been linked to multiple cancers of the nasopharynx, non-Hodgkin lymphoma (NHL), and to Hodgkin's disease (HD) *(18,56,77,78)*. Among 1,549 nasopharyngeal cancer patients, a 14-fold increased risk of second head and neck cancers was reported, and 9% of the patients with multiple head and neck cancers had positive EBV infection compared with only 4% among the control group of patients with only a primary head and neck cancer *(79)*.

Impairment of the immune system and thus lack of control of oncogenic viruses greatly elevates the risk of infection-related cancers *(80)*. Patients with human immunodeficiency virus or with organ transplantation exhibited more than a 100-fold increased risk of NHL or Kaposi Sarcoma (KS) and to a lesser extent of HD, cervical, and skin cancer *(80,81)*. Impaired immune function is supposed to explain the reciprocal increased risk multiple malignancies among patients NHL, HD, KS, and melanoma *(1,77,82,83)*.

5.3. Treatment

The etiologic role of cancer treatment for second primary malignancy has been extensively described by others *(23,32,84)*. Chemotherapy, radiotherapy, and hormonal therapy are largely responsible. Acute sequelae are generally associated to chemotherapy, second cancers arising from as little as a few months to 9 years after therapy. As for radio- and hormonal therapy, chronic sequelae usually develop after a longer latency time of 5 to 10 years. The risk may remain elevated for long periods, although a lowering of the observed risk compared with an earlier follow-up period became visible after 25 years, i.e., after the treatment of HD *(85,86)*. Moreover, therapy-related second cancers also may arise due to combination of treatment modalities and genetic predisposition toward cancers or treatment and external factors, such as lifestyle. Although a considerable proportion of second cancers can be therapy related, one should always also consider the benefit of cancer treatment. The problem is that the knowledge

on side effects develops over time, so that there is always a delay of at least a decade, before it affects the existing regimens and contributes to new regimes that combine good survival chances with a lower risk of adverse effect.

5.3.1. Radiotherapy

Initiation and progression to cancer due to radiation dependent on several factors: 1) dose of radiation, 2) sensitivity of the body tissue, 3) exposure field, 4) age at exposure, and 5) interaction with other increasing or decreasing factor(s). Thyroid, breast, and bone marrow are reported as the most radiosensitive tissues *(24,87)*. The effect of radiation is often amplified when such tissue received radiation at an early age. For example, risk of breast cancer after HD is highest for those who were treated before age 30 (RR = 6–8), whereas risk after age 30 is only minimally elevated *(59,78,86)*. Furthermore, the risk of lung cancer among HD survivors was significantly higher among smokers, based on a multiplicative relation between smoking and radiation (**Table 5.4**; *(62,63)*). Finally, radiation also attenuated the risk of sarcoma among retinoblastoma patients *(88)* who were positive for Rb1 mutation (the cumulative risk at 50 years after diagnosis was 51%), whereas for those without the mutation radiotherapy did not significantly affect their risk of second cancer (cumulative risk was 5%) *(89)*.

5.3.2. Chemotherapy

Chemotherapy is commonly related to the increase occurrence of leukemia (**Table 5.5**), and its leukemogenic effect is more potent than radiation. As for solid tumors, firm evidence has been established between the excess risk of bladder cancer among NHL patients treated with cyclophosphamide; three to seven excess cancers in 100 NHL cases treated with moderate to high dose of cyclophosphamide *(90)*. There also might be an association between alkylating agents and bone sarcoma *(91)* and lung cancer *(92)*, although not all could prove a dose–response (chemotherapy cycle) relationship *(93)*. Furthermore, combination of chemotherapy with radiation may cause a higher risk of a second cancer than for the individual therapy alone *(93–95)*. Lung cancer risk after chemo- and radiotherapy among HD patients was as expected as excess risks were added together. Conversely, the combination with chemotherapy reduced the increased breast cancer risk among patients with HD who received radiotherapy (RR = 3.2 with radiotherapy; RR = 0.6 with chemotherapy; RR = 1.4 with radio- and chemotherapy) *(94)*. Similarly breast cancer patients who received chemotherapy in combination with radiation exhibited half of the second breast cancer compared with those who only received radiotherapy *(46,96)*. Chemotherapy induces early menopause, thus substantially reducing the risk of breast cancer *(71)*.

Table 5.4
Relative risk of subsequent lung cancer by treatment and smoking habits in 19,046 patients treated for Hodgkin's disease *(63)*

Treatment for Hodgkin's disease	Moderate–heavy smokers	
	No. of lung cancers	RR (95% CI)
No	10	6 (1.9–20.4)
Radiation >5 Gy	20	20.2 (6.8–68)
Chemotherapy	33	16.8 (6.2–53)
Radiotherapy and chemotherapy	24	49.1 (15.1 –187)

Reference group was patients without radiation or chemotherapy who were non- or light smokers 5 years before lung cancer diagnosis. Moderate represents individuals who smoked one to two packs a day, and heavy represents individuals who smoked two or more packs a day. Adapted from Travis, L. B., Gospodarowicz, M., Curtis, R. E., Clarke, E. A., Andersson, M., Glimelius, B., Joensuu, T., Lynch, C. F., van Leeuwen, F. E., Holowaty, E., Storm, H., Glimelius, I., Pukkala, E., Stovall, M., Fraumeni, J. F., Jr., Boice, J. D., Jr., Gilbert, E. (2002) Lung cancer following chemotherapy and radiotherapy for Hodgkin's disease. J Natl Cancer Inst. **94,** 182–92. With permission from Oxford University Press.

5.3.3. Hormonal Therapy

Tamoxifen as breast cancer treatment has been consistently related to an elevated risk of endometrial cancer *(72,97,98)*. Endometrial cancer risk increased by two-fold among 2-year tamoxifen users and by 4- to 8-fold among five- or more-year users. Most studies have found no difference in the effect of tamoxifen on endometrial cancer risk between HRT users and nonusers or between obese and nonobese groups *(68,72,99)*, although one study stated otherwise *(100)*.

Conversely, hormonal therapy has demonstrated a protective effect against the normally increased risk of a second primary breast cancer among breast cancer patients *(46,97)*. In the trials where patients took 1, 2, or about 5 years of adjuvant tamoxifen, the 10-year proportional reductions in contralateral breast cancer were 13% (SD = 13), 26% (SD = 9), and 47% (SD = 9), respectively *(101)*.

6. Conclusions and Future Research

Improvements in early detection, diagnosis, and treatment of cancers have increased survival of patients with many types of cancer, however, also carrying a significant increase in number of individuals with multiple malignancies. This problem is larger and will grow even larger in industrialized societies with increasing proportion of elderly persons. Study of occurrence and course of multiple malignancies will improve our insight in etiology and the genesis

Table 5.5
Chemotherapy and related multiple cancers[a]

Chemotherapeutic agents	Treatment for primary cancer	Therapy-related cancer
Alkylating agent (mechlorethamine, chlorambucil, cyclophosphamide, melphalan, semustine, lomustine, carmustine prednimustine, busulfan and dihydroxybusulfan	Lymphomas *(108)* Breast *(13,109)*	Leukemia[b,c]
Platinating agents (cisplatin and carpboplatin)	Ovary *(110,111)* Testis *(112,113)*	Leukemia[b]
Topoisomerase II inhibitors (epipodophyllotoxins etoposide and teniposide)	Lung, testis, solid *(114)* and nonsolid childhood cancers *(115)*	Leukemia[b,d]
Intercalating topoisomerase II inhibitors (anthracycline, doxorubicin and 4-epidoxorubicin)	Lymphomas *(108)* Breast *(13,109)*	Leukemia[b,d]
Cyclophosphamide	NHL *(90)* Ovarian *(116)*	Bladder cancer
Alkylating agent MOPP regimen (mechloretamine, vincristine, procarbazine, prednisone) CHOP regimen (cyclophosphamide, doxorubicin, vincristine, prednisone)	HD *(59,62,63)* NHL *(92)*	Lung cancer
Alkylating agent and anthracycline	Childhood cancers *(91,117)*	Bone sarcoma

[a]Risk of second cancers due to a specific chemotherapeutic agent is difficult to separate because the common combination of several agents. Comprehensive review on the effect of these agents can be found in **ref**. *23*. [b]Leukemia usually implies to acute myeloid leukemia, so far only chronic lymphocytic leukemia has not been linked to chemotherapy *(23)*. [c]More than 50% preceded by myelodysplastic syndrome (MDS). Peak 5–10 years after start of chemotherapy. [d]Not preceded by MDS, peak 2–3 years after the start of chemotherapy.

of cancer in general. Furthermore, human behavior is and will be constantly changing, cancer treatment continuously improving either or not through emerging new therapies; therefore, the need for continuous surveillance. Their effect on the risk of multiple cancers needs a certain period of time until it surfaces and it is clarified unequivocally. Population-based data can serve very well for early warning and verification of "loose" notifications.

The agenda of guideline development for and research of multiple cancers might cover the following six major areas *(34)*, including additional issues beyond what have been laid-out in this chapter:

1. Development of a(n) (inter)national research infrastructure for studies of cancer survivorship; similar registration method is needed to facilitate international collaborative efforts *(15–18,55)*.

2. Creation of a coordinated system of tumor banking for biospecimen collection *(102)*.
3. Development of new technology, bioinformatics, and biomarkers to assess risk and etiologic pathway of multiple cancers, e.g. distinguishing second primary cancers from recurrent or metastatic lesions is important to determine therapy; advances in tumour molecular analysis will certainly improve classification *(103–105)*.
4. Design of new epidemiologic methods and studies; accurate projections of new primary cancer risk among cancers survivors are important to facilitate the surveillance recommendations and are now available for second cancers after childhood leukemia and HD *(106,107)*.
5. Clinical studies assessing the impact of early detection and treatment of a second or higher order malignancy on cause of death and patients' survival and quality of life.
6. Development of evidence-based clinical practice guidelines, including intervention strategies such as behavior modification to prevent occurrence of a new primary cancer; follow-up of cancer survivors; and tailored-therapy for the second, third, or higher order primary cancer, which usually involved the elderly who are more fragile to cancer treatment.

References

1. Fraumeni, J. F. J., Curtis, R., Edwards, B. K., Tucker, M. A. (2006) Introduction. Bethesda, MD: National Cancer Institute. pp. 1–7.
2. Sankila, R., Pukkala, E., Teppo, L. (1995) Risk of subsequent malignant neoplasms among 470,000 cancer patients in Finland, 1953–1991. Int J Cancer **60,** 464–70.
3. Mariotto, A. B., Rowland, J. H., Ries, L. A., Scoppa, S., Feuer, E. J. (2007) Multiple cancer prevalence: a growing challenge in long-term survivorship. Cancer Epidemiol Biomark Prev **16,** 566–71.
4. Parkin, D. M., Bray, F., Ferlay, J., Pisani, P. (2005) Global cancer statistics, 2002. CA Cancer J Clin **55,** 74–108.
5. Janssen-Heijnen, M. L., Houterman, S., Lemmens, V. E., Louwman, M. W., Maas, H. A., Coebergh, J. W. (2005) Prognostic impact of increasing age and co-morbidity in cancer patients: a population-based approach. Crit Rev Oncol Hematol **55,** 231–40.
6. Soerjomataram, I., Louwman, M. W., Ribot, J. G., Roukema, J. A., Coebergh, J. W. (2007) An overview of prognostic factors for long-term survivors of breast cancer. Breast Cancer Res Treat **107,** 309–330.
7. Howe, H. L. (2003) A review of the definition for multiple primary cancers in the United States. Workshop proceedings from December 4–6, 2002, in Princeton, New Jersey. Springfield (IL): North American Association of Central Cancer Registries.
8. Johnson, C. H., Adamo, M. (2007) SEER Program Coding and Staging Manual 2007. Bethesda, MD: National Cancer Institute.
9. Working Group Report. (2005) International rules for multiple primary cancers (ICD-0, third edition). Eur J Cancer Prev **14,** 307–8.
10. Vriesema, J. L., Aben, K. K., Witjes, J. A., Kiemeney, L. A., Schalken, J. A. (2001) Superficial and metachronous invasive bladder carcinomas are clonally related. Int J Cancer **93,** 699–702.
11. Aben, K. K., Witjes, J. A., van Dijck, J. A., Schalken, J. A., Verbeek, A. L., Kiemeney, L. A. (2000) Lower incidence of urothelial

cell carcinoma due to the concept of a clonal origin. Eur J Cancer **36,** 2385–9.

12. Harari, P. M. (1998) Re: Distinguishing second primary tumors from lung metastases in patients with head and neck squamous cell carcinoma. J Natl Cancer Inst **90,** 1571.
13. Smith, R. E., Bryant, J., DeCillis, A., Anderson, S. (2003) Acute myeloid leukemia and myelodysplastic syndrome after doxorubicin-cyclophosphamide adjuvant therapy for operable breast cancer: the National Surgical Adjuvant Breast and Bowel Project Experience. J Clin Oncol **21,** 1195–204.
14. Huang, E., Buchholz, T. A., Meric, F., Krishnamurthy, S., Mirza, N. Q., Ames, F. C., Feig, B. W., Kuerer, H. M., Ross, M. I., Singletary, S. E., McNeese, M. D., Strom, E. A., Hunt, K. K. (2002) Classifying local disease recurrences after breast conservation therapy based on location and histology: new primary tumors have more favorable outcomes than true local disease recurrences. Cancer **95,** 2059–67.
15. Crocetti, E., Lecker, S., Buiatti, E., Storm, H. H. (1996) Problems related to the coding of multiple primary cancers. Eur J Cancer **32A,** 1366–70.
16. Hotes, J. L., Ellison, L. F., Howe, H. L., Friesen, I., Kohler, B. (2004) Variation in breast cancer counts using SEER and IARC multiple primary coding rules. Cancer Causes Control **15,** 185–91.
17. Scelo, G., Boffetta, P., Autier, P., Hemminki, K., Pukkala, E., Olsen, J. H., Weiderpass, E., Tracey, E., Brewster, D. H., McBride, M. L., Kliewer, E. V., Tonita, J. M., Pompe-Kirn, V., Chia, K. S., Jonasson, J. G., Martos, C., Giblin, M., Brennan, P. (2007) Associations between ocular melanoma and other primary cancers: an international population-based study. Int J Cancer **120,** 152–9.
18. Scelo, G., Boffetta, P., Corbex, M., Chia, K. S., Hemminki, K., Friis, S., Pukkala, E., Weiderpass, E., McBride, M. L., Tracey, E., Brewster, D. H., Pompe-Kirn, V., Kliewer, E. V., Tonita, J. M., Martos, C., Jonasson, J. G., Brennan, P. (2007) Second primary cancers in patients with nasopharyngeal carcinoma: a pooled analysis of 13 cancer registries. Cancer Causes Control **18,** 269–78.
19. Mellemkjaer, L., Friis, S., Olsen, J. H., Scelo, G., Hemminki, K., Tracey, E., Andersen, A., Brewster, D. H., Pukkala, E., McBride, M. L., Kliewer, E. V., Tonita, J. M., Kee-Seng, C., Pompe-Kirn, V., Martos, C., Jonasson, J. G., Boffetta, P., Brennan, P. (2006) Risk of second cancer among women with breast cancer. Int J Cancer **118,** 2285–92.
20. Tuohimaa, P., Pukkala, E., Scelo, G., Olsen, J. H., Brewster, D. H., Hemminki, K., Tracey, E., Weiderpass, E., Kliewer, E. V., Pompe-Kirn, V., McBride, M. L., Martos, C., Chia, K. S., Tonita, J. M., Jonasson, J. G., Boffetta, P., Brennan, P. (2007) Does solar exposure, as indicated by the non-melanoma skin cancers, protect from solid cancers: vitamin D as a possible explanation. Eur J Cancer **43,** 1701–1712.
21. Grant, W. B. (2007) A meta-analysis of second cancers after a diagnosis of non-melanoma skin cancer: additional evidence that solar ultraviolet-B irradiance reduces the risk of internal cancers. J Steroid Biochem Mol Biol **103,** 668–74.
22. de Vries, E., Soerjomataram, I., Houterman, S., Louwman, M. W., Coebergh, J. W. (2007) Decreased risk of prostate cancer after skin cancer diagnosis: a protective role of ultraviolet radiation? Am J Epidemiol **165,** 966–72.
23. van Leeuwen, F. E., Travis, L. B. (2005) Second Cancers, 7 ed. Philadelphia. PA: Lippincott Williams and Wilkins. pp. 2575–602.
24. Travis, L. B. (2006) The epidemiology of second primary cancers. Cancer Epidemiol Biomarkers Prev **15,** 2020–6.
25. Curtis, R., Ries, L. A. G. (2006) Methods. Bethesda, MD: National Cancer Institute. pp. 1–7.
26. Schoenberg, B. S., Myers, M. H. (1977) Statistical methods for studying multiple primary malignant neoplasms. Cancer **40,** 1892–8.
27. Begg, C. B., Zhang, Z. F., Sun, M., Herr, H. W., Schantz, S. P. (1995) Methodology for evaluating the incidence of second primary cancers with application to smoking-related cancers from the Surveillance, Epidemiology, and End Results (SEER) program. Am J Epidemiol **142,** 653–65.
28. Cutler, S. J., Ederer, F. (1958) Maximum utilization of the life table method in analyzing survival. J Chronic Dis **8,** 699–712.
29. Gooley, T. A., Leisenring, W., Crowley, J., Storer, B. E. (1999) Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. Stat Med **18,** 695–706.
30. Yu, G. P., Schantz, S. P., Neugut, A. I., Zhang, Z. F. (2006) Incidences and trends of second cancers in female breast cancer patients: a fixed inception cohort-based analysis (United States). Cancer Causes Control **17,** 411–20.

31. Yasui, Y., Liu, Y., Neglia, J. P., Friedman, D. L., Bhatia, S., Meadows, A. T., Diller, L. R., Mertens, A. C., Whitton, J., Robison, L. L. (2003) A methodological issue in the analysis of second-primary cancer incidence in long-term survivors of childhood cancers. Am J Epidemiol **158,** 1108–13.
32. Travis, L. B. (2002) Therapy-associated solid tumors. Acta Oncol. **41,** 323–33.
33. Largent, J. A., Capanu, M., Bernstein, L., Langholz, B., Mellemkjaer, L., Malone, K. E., Begg, C. B., Haile, R. W., Lynch, C. F., Anton-Culver, H., Wolitzer, A., Bernstein, J. L. (2007) Reproductive History and Risk of Second Primary Breast Cancer: the WECARE Study. Cancer Epidemiol Biomarkers Prev **16,** 906–911.
34. Travis, L. B., Rabkin, C. S., Brown, L. M., Allan, J. M., Alter, B. P., Ambrosone, C. B., Begg, C. B., Caporaso, N., Chanock, S., DeMichele, A., Figg, W. D., Gospodarowicz, M. K., Hall, E. J., Hisada, M., Inskip, P., Kleinerman, R., Little, J. B., Malkin, D., Ng, A. K., Offit, K., Pui, C. H., Robison, L. L., Rothman, N., Shields, P. G., Strong, L., Taniguchi, T., Tucker, M. A., Greene, M. H. (2006) Cancer survivorship--genetic susceptibility and second primary cancers: research strategies and recommendations. J Natl Cancer Inst **98,** 15–25.
35. Garber, J. E., Offit, K. (2005) Hereditary cancer predisposition syndromes. J Clin Oncol **23,** 276–92.
36. Fearon, E. R. (1997) Human cancer syndromes: clues to the origin and nature of cancer. Science **278,** 1043–50.
37. Nagy, R., Sweet, K., Eng, C. (2004) Highly penetrant hereditary cancer syndromes. Oncogene **23,** 6445–70.
38. Hill, D. A., Gilbert, E., Dores, G. M., Gospodarowicz, M., van Leeuwen, F. E., Holowaty, E., Glimelius, B., Andersson, M., Wiklund, T., Lynch, C. F., Van't Veer, M., Storm, H., Pukkala, E., Stovall, M., Curtis, R. E., Allan, J. M., Boice, J. D., Travis, L. B. (2005) Breast cancer risk following radiotherapy for Hodgkin lymphoma: modification by other risk factors. Blood **106,** 3358–65.
39. Landgren, O., Pfeiffer, R. M., Stewart, L., Gridley, G., Mellemkjaer, L., Hemminki, K., Goldin, L. R., Travis, L. B. (2007) Risk of second malignant neoplasms among lymphoma patients with a family history of cancer. Int J Cancer **120,** 1099–102.
40. Trentham-Dietz, A., Newcomb, P. A., Nichols, H. B., Hampton, J. M. (2007) Breast cancer risk factors and second primary malignancies among women with breast cancer. Breast Cancer Res Treat **105,** 359–68.
41. Kmet, L. M., Cook, L. S., Weiss, N. S., Schwartz, S. M., White, E. (2003) Risk factors for colorectal cancer following breast cancer. Breast Cancer Res Treat **79,** 143–7.
42. Hisada, M., Garber, J. E., Fung, C. Y., Fraumeni, J. F., Jr., Li, F. P. (1998) Multiple primary cancers in families with Li-Fraumeni syndrome. J Natl Cancer Inst **90,** 606–11.
43. Hemminki, K., Ji, J., Forsti, A. (2007) Risks for familial and contralateral breast cancer interact multiplicatively and cause a high risk. Cancer Res **67,** 868–70.
44. Hemminki, K., Li, X. (2005) Familial risk for lung cancer by histology and age of onset: evidence for recessive inheritance. Exp Lung Res. **31,** 205–15.
45. Soerjomataram, I., Louwman, W. J., de Vries, E., Lemmens, V. E., Klokman, W. J., Coebergh, J. W. (2005) Primary malignancy after primary female breast cancer in the South of the Netherlands, 1972–2001. Breast Cancer Res Treat **93,** 91–5.
46. Soerjomataram, I., Louwman, W. J., Lemmens, V. E., de Vries, E., Klokman, W. J., Coebergh, J. W. (2005) Risks of second primary breast and urogenital cancer following female breast cancer in the south of The Netherlands, 1972–2001. Eur J Cancer **41,** 2331–7.
47. Raymond, J. S., Hogue, C. J. (2006) Multiple primary tumours in women following breast cancer, 1973–2000. Br J Cancer **94,** 1745–50.
48. Metcalfe, K. A., Lynch, H. T., Ghadirian, P., Tung, N., Olivotto, I. A., Foulkes, W. D., Warner, E., Olopade, O., Eisen, A., Weber, B., McLennan, J., Sun, P., Narod, S. A. (2005) The risk of ovarian cancer after breast cancer in BRCA1 and BRCA2 carriers. Gynecol Oncol **96,** 222–6.
49. Metcalfe, K., Lynch, H. T., Ghadirian, P., Tung, N., Olivotto, I., Warner, E., Olopade, O. I., Eisen, A., Weber, B., McLennan, J., Sun, P., Foulkes, W. D., Narod, S. A. (2004) Contralateral breast cancer in BRCA1 and BRCA2 mutation carriers. J Clin Oncol **22,** 2328–35.
50. Lynch, H. T., Shaw, T. G., Lynch, J. F. (2004) Inherited predisposition to cancer: a historical overview. Am J Med Genet C Semin Med Genet **129,** 5–22.
51. Meijers-Heijboer, H., Wijnen, J., Vasen, H., Wasielewski, M., Wagner, A., Hollestelle, A., Elstrodt, F., van den Bos, R., de Snoo,

A., Fat, G. T., Brekelmans, C., Jagmohan, S., Franken, P., Verkuijlen, P., van den Ouweland, A., Chapman, P., Tops, C., Moslein, G., Burn, J., Lynch, H., Klijn, J., Fodde, R., Schutte, M. (2003) The CHEK2 1100delC mutation identifies families with a hereditary breast and colorectal cancer phenotype. Am J Hum Genet **72,** 1308–14.

52. Dong, C., Hemminki, K. (2001) Multiple primary cancers of the colon, breast and skin (melanoma) as models for polygenic cancers. Int J Cancer. **92,** 883–7.
53. Balmain, A., Gray, J., Ponder, B. (2003) The genetics and genomics of cancer. Nat Genet 33 Suppl, 238–44.
54. Levi, F., Randimbison, L., Maspoli, M., Te, V. C., La Vecchia, C. (2007) Second neoplasms after oesophageal cancer. Int J Cancer **121,** 694–7.
55. Shen, M., Boffetta, P., Olsen, J. H., Andersen, A., Hemminki, K., Pukkala, E., Tracey, E., Brewster, D. H., McBride, M. L., Pompe-Kirn, V., Kliewer, E. V., Tonita, J. M., Chia, K. S., Martos, C., Jonasson, J. G., Colin, D., Scelo, G., Brennan, P. (2006) A pooled analysis of second primary pancreatic cancer. Am J Epidemiol **163,** 502–11.
56. Brown, L. M., McCarron, P., Freedman, D. M. (2006) New malignancies following cancer of the buccal cavity and pharynx. Bethesda, MD: National Cancer Institute. pp. 1–7.
57. Caporaso, N., Dodd, K. W., Tucker, M. A. (2006) New malignancies following cancer of the respiratory tract. Bethesda, MD: National Cancer Institute. 1–7.
58. Dikshit, R. P., Boffetta, P., Bouchardy, C., Merletti, F., Crosignani, P., Cuchi, T., Ardanaz, E., Brennan, P. (2005) Risk factors for the development of second primary tumors among men after laryngeal and hypopharyngeal carcinoma. Cancer **103,** 2326–33.
59. Swerdlow, A. J., Barber, J. A., Hudson, G. V., Cunningham, D., Gupta, R. K., Hancock, B. W., Horwich, A., Lister, T. A., Linch, D. C. (2000) Risk of second malignancy after Hodgkin's disease in a collaborative British cohort: the relation to age at treatment. J Clin Oncol **18,** 498–509.
60. Tucker, M. A., Murray, N., Shaw, E. G., Ettinger, D. S., Mabry, M., Huber, M. H., Feld, R., Shepherd, F. A., Johnson, D. H., Grant, S. C., Aisner, J., Johnson, B. E. (1997) Second primary cancers related to smoking and treatment of small-cell lung cancer. Lung Cancer Working Cadre. J Natl Cancer Inst **89,** 1782–8.
61. Khuri, F. R., Kim, E. S., Lee, J. J., Winn, R. J., Benner, S. E., Lippman, S. M., Fu, K. K., Cooper, J. S., Vokes, E. E., Chamberlain, R. M., Williams, B., Pajak, T. F., Goepfert, H., Hong, W. K. (2001) The impact of smoking status, disease stage, and index tumor site on second primary tumor incidence and tumor recurrence in the head and neck retinoid chemoprevention trial. Cancer Epidemiol Biomark Prev **10,** 823–9.
62. Gilbert, E. S., Stovall, M., Gospodarowicz, M., Van Leeuwen, F. E., Andersson, M., Glimelius, B., Joensuu, T., Lynch, C. F., Curtis, R. E., Holowaty, E., Storm, H., Pukkala, E., van't Veer, M. B., Fraumeni, J. F., Boice, J. D., Jr., Clarke, E. A., Travis, L. B. (2003) Lung cancer after treatment for Hodgkin's disease: focus on radiation effects. Radiat Res **159,** 161–73.
63. Travis, L. B., Gospodarowicz, M., Curtis, R. E., Clarke, E. A., Andersson, M., Glimelius, B., Joensuu, T., Lynch, C. F., van Leeuwen, F. E., Holowaty, E., Storm, H., Glimelius, I., Pukkala, E., Stovall, M., Fraumeni, J. F., Jr., Boice, J. D., Jr., Gilbert, E. (2002) Lung cancer following chemotherapy and radiotherapy for Hodgkin's disease. J Natl Cancer Inst **94,** 182–92.
64. Do, K. A., Johnson, M. M., Lee, J. J., Wu, X. F., Dong, Q., Hong, W. K., Khuri, F. R., Spitz, M. R. (2004) Longitudinal study of smoking patterns in relation to the development of smoking-related secondary primary tumors in patients with upper aerodigestive tract malignancies. Cancer **101,** 2837–42.
65. Pike, M. C., Pearce, C. L., Wu, A. H. (2004) Prevention of cancers of the breast, endometrium and ovary. Oncogene **23,** 6379–91.
66. Samanic, C., Chow, W. H., Gridley, G., Jarvholm, B., Fraumeni, J. F., Jr. (2006) Relation of body mass index to cancer risk in 362,552 Swedish men. Cancer Causes Control **17,** 901–9.
67. Mysliwiec, P. A., Cronin, K. A., Schatzkin, A. (2006) New Malignancies following cancer of the colon, rectum, and Anus. Bethesda, MD: National Cancer Institute. 1–7.
68. Dignam, J. J., Wieand, K., Johnson, K. A., Fisher, B., Xu, L., Mamounas, E. P. (2003) Obesity, tamoxifen use, and outcomes in women with estrogen receptor-positive early-stage breast cancer. J Natl Cancer Inst **95,** 1467–76.
69. Dignam, J. J., Wieand, K., Johnson, K. A., Raich, P., Anderson, S. J., Somkin, C., Wickerham, D. L. (2006) Effects of obesity and race on prognosis in lymph node-nega-

tive, estrogen receptor-negative breast cancer. Breast Cancer Res Treat **97,** 245–54.

70. Dignam, J. J., Polite, B. N., Yothers, G., Raich, P., Colangelo, L., O'Connell, M. J., Wolmark, N. (2006) Body mass index and outcomes in patients who receive adjuvant chemotherapy for colon cancer. J Natl Cancer Inst **98,** 1647–54.
71. van Leeuwen, F. E., Klokman, W. J., Stovall, M., Dahler, E. C., van't Veer, M. B., Noordijk, E. M., Crommelin, M. A., Aleman, B. M., Broeks, A., Gospodarowicz, M., Travis, L. B., Russell, N. S. (2003) Roles of radiation dose, chemotherapy, and hormonal factors in breast cancer following Hodgkin's disease. J Natl Cancer Inst **95,** 971–80.
72. Swerdlow, A. J., Jones, M. E. (2005) Tamoxifen treatment for breast cancer and risk of endometrial cancer: a case-control study. J Natl Cancer Inst **97,** 375-84.
73. Kleinerman, R. A., Kosary, C., Hildesheim, A. (2006) New malignancies following cancer of the cervix uteri, vagina, and vulva. Bethesda, MD: National Cancer Institute. 1–7.
74. Hemminki, K., Dong, C., Vaittinen, P. (2000) Second primary cancer after in situ and invasive cervical cancer. Epidemiology **11,** 457–61.
75. Levi, F., Te, V. C., Randimbison, L., La Vecchia, C. (2001) Second primary cancers after in situ and invasive cervical cancer. Epidemiology **12,** 281–2.
76. Vinokurova, S., Wentzensen, N., Einenkel, J., Klaes, R., Ziegert, C., Melsheimer, P., Sartor, H., Horn, L. C., Hockel, M., von Knebel Doeberitz, M. (2005) Clonal history of papillomavirus-induced dysplasia in the female lower genital tract. J Natl Cancer Inst **97,** 1816–21.
77. Brennan, P., Scelo, G., Hemminki, K., Mellemkjaer, L., Tracey, E., Andersen, A., Brewster, D. H., Pukkala, E., McBride, M. L., Kliewer, E. V., Tonita, J. M., Seow, A., Pompe-Kirn, V., Martos, C., Jonasson, J. G., Colin, D., Boffetta, P. (2005) Second primary cancers among 109 000 cases of non-Hodgkin's lymphoma. Br J Cancer **93,** 159–66.
78. Dores, G. M., Cote, T. R., Travis, L. B. (2006) New malignancies following Hodgkin lymphoma, Non-Hodgkin Lymphoma, and Myeloma. Bethesda, MD: National Cancer Institute. 1–7.
79. Wang, C. C., Chen, M. L., Hsu, K. H., Lee, S. P., Chen, T. C., Chang, Y. S., Tsang, N. M., Hong, J. H. (2000) Second malignant tumors in patients with nasopharyngeal carcinoma and their association with Epstein-Barr virus. Int J Cancer **87,** 228–31.
80. Schwartz, R. S. (2001) Immunodeficiency, immunosuppression, and susceptibility to neoplasms. J Natl Cancer Inst Monogr 5–9.
81. Clifford, G. M., Polesel, J., Rickenbach, M., Dal Maso, L., Keiser, O., Kofler, A., Rapiti, E., Levi, F., Jundt, G., Fisch, T., Bordoni, A., De Weck, D., Franceschi, S. (2005) Cancer risk in the Swiss HIV Cohort Study: associations with immunodeficiency, smoking, and highly active antiretroviral therapy. J Natl Cancer Inst **97,** 425–32.
82. Hemminki, K., Jiang, Y., Steineck, G. (2003) Skin cancer and non-Hodgkin's lymphoma as second malignancies. markers of impaired immune function? Eur J Cancer **39,** 223–9.
83. Metayer, C., Lynch, C. F., Clarke, E. A., Glimelius, B., Storm, H., Pukkala, E., Joensuu, T., van Leeuwen, F. E., van't Veer, M. B., Curtis, R. E., Holowaty, E. J., Andersson, M., Wiklund, T., Gospodarowicz, M., Travis, L. B. (2000) Second cancers among long-term survivors of Hodgkin's disease diagnosed in childhood and adolescence. J Clin Oncol **18,** 2435–43.
84. Allan, J. M., Travis, L. B. (2005) Mechanisms of therapy-related carcinogenesis. Nat Rev Cancer **5,** 943–55.
85. Dores, G. M., Metayer, C., Curtis, R. E., Lynch, C. F., Clarke, E. A., Glimelius, B., Storm, H., Pukkala, E., van Leeuwen, F. E., Holowaty, E. J., Andersson, M., Wiklund, T., Joensuu, T., van't Veer, M. B., Stovall, M., Gospodarowicz, M., Travis, L. B. (2002) Second malignant neoplasms among long-term survivors of Hodgkin's disease: a population-based evaluation over 25 years. J Clin Oncol **20,** 3484–94.
86. van Leeuwen, F. E., Klokman, W. J., Veer, M. B., Hagenbeek, A., Krol, A. D., Vetter, U. A., Schaapveld, M., van Heerde, P., Burgers, J. M., Somers, R., Aleman, B. M. (2000) Long-term risk of second malignancy in survivors of Hodgkin's disease treated during adolescence or young adulthood. J Clin Oncol **18,** 487–97.
87. Boice, J. D. Ionizing radiation. 3 ed. New York: Oxford University Press. 259–93.
88. Kleinerman, R. A., Tucker, M. A., Abramson, D. H., Seddon, J. M., Tarone, R. E., Fraumeni, J. F., Jr. (2007) Risk of soft tissue sarcomas by individual subtype in survivors of hereditary retinoblastoma. J Natl Cancer Inst **99,** 24–31.
89. Wong, F. L., Boice, J. D., Jr., Abramson, D. H., Tarone, R. E., Kleinerman, R. A., Stovall, M., Goldman, M. B., Seddon, J. M., Tarbell, N., Fraumeni, J. F., Jr., Li, F.

P. (1997) Cancer incidence after retinoblastoma. Radiation dose and sarcoma risk. JAMA **278,** 1262–7.

90. Travis, L. B., Curtis, R. E., Glimelius, B., Holowaty, E. J., Van Leeuwen, F. E., Lynch, C. F., Hagenbeek, A., Stovall, M., Banks, P. M., Adami, J., et al. (1995) Bladder and kidney cancer following cyclophosphamide therapy for non-Hodgkin's lymphoma. J Natl Cancer Inst **87,** 524–30.
91. Henderson, T. O., Whitton, J., Stovall, M., Mertens, A. C., Mitby, P., Friedman, D., Strong, L. C., Hammond, S., Neglia, J. P., Meadows, A. T., Robison, L., Diller, L. (2007) Secondary sarcomas in childhood cancer survivors: a report from the Childhood Cancer Survivor Study. J Natl Cancer Inst **99,** 300–8.
92. Mudie, N. Y., Swerdlow, A. J., Higgins, C. D., Smith, P., Qiao, Z., Hancock, B. W., Hoskin, P. J., Linch, D. C. (2006) Risk of second malignancy after non-Hodgkin's lymphoma: a British Cohort Study. J Clin Oncol **24,** 1568–74.
93. Garwicz, S., Anderson, H., Olsen, J. H., Dollner, H., Hertz, H., Jonmundsson, G., Langmark, F., Lanning, M., Moller, T., Sankila, R., Tulinius, H. (2000) Second malignant neoplasms after cancer in childhood and adolescence: a population-based case-control study in the 5 Nordic countries. The Nordic Society for Pediatric Hematology and Oncology. The Association of the Nordic Cancer Registries. Int J Cancer **88,** 672–8.
94. Travis, L. B., Hill, D. A., Dores, G. M., Gospodarowicz, M., van Leeuwen, F. E., Holowaty, E., Glimelius, B., Andersson, M., Wiklund, T., Lynch, C. F., Van't Veer, M. B., Glimelius, I., Storm, H., Pukkala, E., Stovall, M., Curtis, R., Boice, J. D., Jr., Gilbert, E. (2003) Breast cancer following radiotherapy and chemotherapy among young women with Hodgkin disease. JAMA **290,** 465–75.
95. Guerin, S., Guibout, C., Shamsaldin, A., Dondon, M. G., Diallo, I., Hawkins, M., Oberlin, O., Hartmann, O., Michon, J., Le Deley, M. C., de Vathaire, F. (2007) Concomitant chemo-radiotherapy and local dose of radiation as risk factors for second malignant neoplasms after solid cancer in childhood: a case-control study. Int J Cancer **120,** 96–102.
96. Rubino, C., de Vathaire, F., Shamsaldin, A., Labbe, M., Le, M. G. (2003) Radiation dose, chemotherapy, hormonal treatment and risk of second cancer after breast cancer treatment. Br J Cancer **89,** 840–6.
97. (2005) Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet **365,** 1687–717.
98. Fisher, B., Costantino, J. P., Wickerham, D. L., Redmond, C. K., Kavanah, M., Cronin, W. M., Vogel, V., Robidoux, A., Dimitrov, N., Atkins, J., Daly, M., Wieand, S., Tan-Chiu, E., Ford, L., Wolmark, N. (1998) Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. J Natl Cancer Inst **90,** 1371–88.
99. Bergman, L., Beelen, M. L., Gallee, M. P., Hollema, H., Benraadt, J., van Leeuwen, F. E. (2000) Risk and prognosis of endometrial cancer after tamoxifen for breast cancer. Comprehensive Cancer Centres' ALERT Group. Assessment of Liver and Endometrial cancer Risk following Tamoxifen. Lancet **356,** 881–7.
100. Bernstein, J. L., Thompson, W. D., Risch, N., Holford, T. R. (1992) Risk factors predicting the incidence of second primary breast cancer among women diagnosed with a first primary breast cancer. Am J Epidemiol **136,** 925–36.
101. (1998) Tamoxifen for early breast cancer: an overview of the randomised trials. Early Breast Cancer Trialists' Collaborative Group. Lancet **351,** 1451–67.
102. Teodorovic, I., Isabelle, M., Carbone, A., Passioukov, A., Lejeune, S., Jamine, D., Therasse, P., Gloghini, A., Dinjens, W. N., Lam, K. H., Oomen, M. H., Spatz, A., Ratcliffe, C., Knox, K., Mager, R., Kerr, D., Pezzella, F., van Damme, B., van de Vijver, M., van Boven, H., Morente, M. M., Alonso, S., Kerjaschki, D., Pammer, J., Lopez-Guerrero, J. A., Llombart Bosch, A., van Veen, E. B., Oosterhuis, J. W., Riegman, P. H. (2006) TuBaFrost 6: virtual microscopy in virtual tumour banking. Eur J Cancer **42,** 3110–6.
103. Leong, P. P., Rezai, B., Koch, W. M., Reed, A., Eisele, D., Lee, D. J., Sidransky, D., Jen, J., Westra, W. H. (1998) Distinguishing second primary tumors from lung metastases in patients with head and neck squamous cell carcinoma. J Natl Cancer Inst **90,** 972–7.
104. Blokx, W. A., Lesterhuis, J. J., Andriessen, M. P., Verdijk, M. A., Punt, K. J., Ligtenberg, M. J. (2007) CDKN2A (INK4A-ARF) mutation analysis to distinguish cutaneous melanoma metastasis from a second primary melanoma. Am J Surg Pathol **31,** 637–41.

105. Geurts, T. W., Nederlof, P. M., van den Brekel, M. W., van't Veer, L. J., de Jong, D., Hart, A. A., van Zandwijk, N., Klomp, H., Balm, A. J., van Velthuysen, M. L. (2005) Pulmonary squamous cell carcinoma following head and neck squamous cell carcinoma: metastasis or second primary? Clin Cancer Res **11,** 6608–14.

106. Travis, L. B., Hill, D., Dores, G. M., Gospodarowicz, M., van Leeuwen, F. E., Holowaty, E., Glimelius, B., Andersson, M., Pukkala, E., Lynch, C. F., Pee, D., Smith, S. A., Van't Veer, M. B., Joensuu, T., Storm, H., Stovall, M., Boice, J. D., Jr., Gilbert, E., Gail, M. H. (2005) Cumulative absolute breast cancer risk for young women treated for Hodgkin lymphoma. J Natl Cancer Inst **97,** 1428–37.

107. Hijiya, N., Hudson, M. M., Lensing, S., Zacher, M., Onciu, M., Behm, F. G., Razzouk, B. I., Ribeiro, R. C., Rubnitz, J. E., Sandlund, J. T., Rivera, G. K., Evans, W. E., Relling, M. V., Pui, C. H. (2007) Cumulative incidence of secondary neoplasms as a first event after childhood acute lymphoblastic leukemia. JAMA **297,** 1207–15.

108. Schonfeld, S. J., Gilbert, E. S., Dores, G. M., Lynch, C. F., Hodgson, D. C., Hall, P., Storm, H., Andersen, A., Pukkala, E., Holowaty, E., Kaijser, M., Andersson, M., Joensuu, H., Fossa, S. D., Allan, J. M., Travis, L. B. (2006) Acute myeloid leukemia following Hodgkin lymphoma: a population-based study of 35,511 patients. J Natl Cancer Inst **98,** 215–8.

109. Howard, R. A., Gilbert, E. S., Chen, B. E., Hall, P., Storm, H., Pukkala, E., Langmark, F., Kaijser, M., Andersson, M., Joensuu, H., Fossa, S. D., Travis, L. B. (2007) Leukemia following breast cancer: an international population-based study of 376,825 women. Breast Cancer Res Treat **105,** 359–368.

110. Travis, L. B., Holowaty, E. J., Bergfeldt, K., Lynch, C. F., Kohler, B. A., Wiklund, T., Curtis, R. E., Hall, P., Andersson, M., Pukkala, E., Sturgeon, J., Stovall, M. (1999) Risk of leukemia after platinum-based chemotherapy for ovarian cancer. N Engl J Med **340,** 351–7.

111. Kaldor, J. M., Day, N. E., Pettersson, F., Clarke, E. A., Pedersen, D., Mehnert, W., Bell, J., Host, H., Prior, P., Karjalainen, S., et al. (1990) Leukemia following chemotherapy for ovarian cancer. N Engl J Med **322,** 1–6.

112. Robinson, D., Moller, H., Horwich, A. (2007) Mortality and incidence of second cancers following treatment for testicular cancer. Br J Cancer **96,** 529–33.

113. Travis, L. B., Andersson, M., Gospodarowicz, M., van Leeuwen, F. E., Bergfeldt, K., Lynch, C. F., Curtis, R. E., Kohler, B. A., Wiklund, T., Storm, H., Holowaty, E., Hall, P., Pukkala, E., Sleijfer, D. T., Clarke, E. A., Boice, J. D., Jr., Stovall, M., Gilbert, E. (2000) Treatment-associated leukemia following testicular cancer. J Natl Cancer Inst **92,** 1165–71.

114. Le Deley, M. C., Leblanc, T., Shamsaldin, A., Raquin, M. A., Lacour, B., Sommelet, D., Chompret, A., Cayuela, J. M., Bayle, C., Bernheim, A., de Vathaire, F., Vassal, G., Hill, C. (2003) Risk of secondary leukemia after a solid tumor in childhood according to the dose of epipodophyllotoxins and anthracyclines: a case-control study by the Societe Francaise d'Oncologie Pediatrique. J Clin Oncol **21,** 1074–81.

115. Neglia, J. P., Friedman, D. L., Yasui, Y., Mertens, A. C., Hammond, S., Stovall, M., Donaldson, S. S., Meadows, A. T., Robison, L. L. (2001) Second malignant neoplasms in five-year survivors of childhood cancer: childhood cancer survivor study. J Natl Cancer Inst **93,** 618–29.

116. Kaldor, J. M., Day, N. E., Kittelmann, B., Pettersson, F., Langmark, F., Pedersen, D., Prior, P., Neal, F., Karjalainen, S., Bell, J., et al. (1995) Bladder tumours following chemotherapy and radiotherapy for ovarian cancer: a case-control study. Int J Cancer **63,** 1–6.

117. Hawkins, M. M., Wilson, L. M., Burton, H. S., Potok, M. H., Winter, D. L., Marsden, H. B., Stovall, M. A. (1996) Radiotherapy, alkylating agents, and risk of bone cancer after childhood cancer. J Natl Cancer Inst **88,** 270–8.

Chapter 6

Cancer Screenings, Diagnostic Technology Evolution, and Cancer Control

Fabrizio Stracci

Summary

Screening should allow for the anticipation of cancer diagnosis at an earlier stage, when curative treatment is possible. Screening for cervical, large bowel, and breast cancer were shown to be effective in reducing mortality. The wide acceptance of the screening concept led to the wide diffusion also of screening of uncertain benefit against prostate cancer and skin melanoma. Diagnostic technologies are continuously evolving, and new tests are proposed to improve existing screenings or as screening tests for additional cancer sites (e.g., lung cancer). Cancer screening, however, is a complex and costly intervention that does not result only in benefits but also may cause harm. A major emerging problem of screening is overdiagnosis, or the detection of cases that would have not progressed to the symptomatic phase in the absence of screening. Thus, both experimental and observational evaluation studies are needed to reduce harm caused by screenings and to select effective interventions among many proposed innovations. Finally, the research of markers to assess the aggressive nature of screen-detected lesions is of great importance to improve screenings' harm/benefit ratio.

Key words: Screening, cancer control, diagnostic test, evaluation, public health, overdiagnosis.

1. Screening Definition

Early diagnosis is the identification of a disease or a predisease status close to the beginning of its natural history. "Early" refers to the successive steps or phases over the disease progression, independently of the time required for progression. Rather, progression time influences the probability of early diagnosis, together with other factors such as precocity of symptoms and accessibility of health services. Early diagnosis of a cancer or a precancerous lesion is desirable whenever it allows for effective treatment and better health outcomes. Active search for a disease in asymptomatic

Mukesh Verma (ed.), *Methods in Molecular Biology, Cancer Epidemiology, vol. 471*
© 2009 Humana Press, a part of Springer Science + Business Media, Totowa, NJ
Book doi: 10.1007/978-1-59745-416-2

individuals (screening) and the level of medical awareness together with accessibility of health services both may influence cancer stage at diagnosis. Moreover the availability of a screening is likely to be associated with population awareness *(1)*.

Early diagnosis allowing for more effective treatment and better health outcomes is the rationale for screening as a cancer control strategy. Asymptomatic people are examined to find out those more likely to have the disease that is the object of screening. The asymptomatic criterion to define screening seems somewhat fuzzy in the case of skin melanoma: the same lesion would be screen detected if detected by the expert look of a dermatologist and symptomatic if noticed by the patient or by a familiar. People testing positive undergo further diagnostic effort to confirm the presence of disease and, eventually are referred for treatment. Thus, screening is the application of one or more diagnostic tests in an asymptomatic person to identify the disease (i.e., cancer or precancerous lesions) at an earlier stage than symptomatic diagnosis. A screening test capable of identifying cancer at an early stage without a more effective treatment does not modify disease progression and thus is not desirable, because the main result would be prolonging time lived as a cancer patient. For a disease to be suitable for screening, there should be a preclinical phase when it is detectable *(2)*. Conceptually, we may divide the preclinical phase into two parts: the effective preclinical phase in which detection at screening leads to treatment conferring a health benefit and an ineffective preclinical phase in which the only consequence of screening is anticipation of diagnosis. The time spent in the preclinical phase is called sojourn time. Sojourn time is not constant; instead, it is a feature of each individual cancer (i.e., each lesion will progress through the preclinical phase at its own speed that may even vary over time). Rapidly progressive lesions have a short sojourn time and a correspondingly short stay in the effective preclinical phase. Conversely, slowly progressive lesions may require a time longer than the individual life span to become symptomatic; such lesions have no effective preclinical phase because there is no benefit to be gained with treatment. The diagnosis of lesions that would not have progressed to clinical diagnosis in the absence of screening is called overdiagnosis. Some screen-detectable lesions may even be able to regress spontaneously (e.g., low-grade cervical dysplasia). The distribution of sojourn times may vary by age. For example, breast cancer cases in young women (<40 years old) were reported to be more aggressive than in older women *(3)*. In principle, we could use repeated mammography screening and accommodate the screening pace by age to catch rapidly progressive lesions. Shrinking the time between successive tests, however, comes at the price of more false

positive cases and false diagnoses, lower adhesion, and increased costs, and thus may be less cost-effective. Moreover, for screening to be cost-effective, disease prevalence in the target population should be high and prevalence often shows large variation by age. Population with a high quote of aggressive cases and low disease prevalence (e.g., breast cancer in women <40 years old) are not amenable to mass screening. However, elderly populations with high prevalence and, perhaps, more slowly progressive cancers but with limited life expectancy may receive less benefit from screening, and they are at higher risk of overdiagnosis. The longer the preclinical phase persists, the greater the opportunity for screening *(4)*, but also the scope for overdiagnosis. Success of a screening program will depend on the quote of progressive lesions detected during the effective preclinical phase and on the number of deaths prevented (related to prevalence); thus, cancer screenings tend to be performed in adults and young old depending on cancer site.

Cancer screening is increasingly being used as a tool for cancer control in terms of both participation into screening and of number of cancer sites with ongoing screening activities. For example, in the United States, 40% of women ~40 years reported a mammography within 2 years in 1987, and this figure was as high as 70% for in 2000 *(5)*. As to cancer sites, screening activities include breast, large bowel, cervix uteri, skin melanoma, and prostate; moreover, interest was expressed and research to evaluate screening opportunities is ongoing for lung, oral, liver, stomach, ovary, neuroblastoma, and urinary bladder screening.

2. Screening Outcomes

Screenings are aimed at reducing mortality. Other major benefits that may be produced by screening include morbidity reduction, longer disease-free survival, and disability reduction (e.g., endoscopic removal of an intestinal polyp instead of surgical resection for invasive cancer). Additional benefits include prolonged survival, less aggressive treatment for early stage disease, and reassurance for true negative screened persons. Major risks of harm of screening includes all health consequences of diagnosis and treatment in persons with lesions that would have not progressed to the symptomatic phase in the absence of screening (overdiagnosis leading to overtreatment) and adverse effect of testing that rarely may be severe (e.g., perforation during colonoscopy). Other risks of harm include consequences of false-positive results (invasive diagnostic testing and psychological consequences) and prolonged time lived with the knowledge of having cancer (corresponding to lead time).

3. Screening and Evidence

Benefits of screening should outweigh harm. However, the evaluation of various benefits and, even more, of harm produced by screenings is a difficult task. Thus, it seems that cancer screening is mainly evaluated in terms of the main health outcome, mortality reduction *(6)*. Randomized controlled trials (well conducted) provide the best evidence of the efficacy of a screening, usually in terms of mortality reduction, because the introduction of a screening intervention produces profound modification of the epidemiology of the disease; thus, evidence from observational studies may suffer from several biases *(7)*. Evidence from randomized trials is presently available for two (large bowel and breast cancer) of five most diffused cancer screenings; trials to evaluate screening for prostate cancer *(8,9)* and melanoma of the skin *(10)* are ongoing, but prostate specific antigen (PSA) testing and skin examination gained widespread diffusion before the availability of the trials' results or even before the trials' design. Evidence of efficacy is deemed sufficient for large bowel fecal occult blood testing (FOBT)-based screening and for pap smear-based testing, although results from randomized studies are available only for large bowel FOBT-based screening. Controversies exist on mammography-based breast cancer screening, despite the availability of trials' results. The efficacy of mammography screening was questioned in a first systematic review by Gotzche and Olsen *(11)*, but two recent systematic reviews confirmed that screening is effective in reducing disease-specific mortality *(12,13)*. However, controversies persist on the quantification of mortality reduction and on the level of overdiagnosis associated with this screening *(14)*.

4. Screening, Evaluation, and Diagnostic Technology Evolution

The use of PSA-based screening is so common that screening diffusion in the control group (contamination) may hamper the interpretation of ongoing trials *(15)*. While for the study of melanoma, dimension was a barrier to the execution of experimental studies, an argument for implementing prostate cancer screening before proof of efficacy was time required for trials' results to be available *(16)*. Because trials may take years to produce results, this study design is increasingly becoming a tool contributing information on efficacy/effectiveness of health intervention used in general practice and not a minimal prerequisite to be produced before implementation *(17)*.

The availability of new technologies is determining the re-evaluation of lung cancer screening; a screening that previously turned out to be ineffective. Lung cancer, due to features such as health burden and advanced stage at diagnosis, is an ideal candidate to screening. Three trials carried out in 1970s did not show any benefit for lung cancer screening based on chest X-ray with or without sputum cytology *(18)*. The case for lung cancer screening also was raised to warn against diffusion of the unproved PSA-based screening for prostate cancer *(19)*. The proposal of lung cancer screening was recently reopened for the availability of the low-dose computed tomography (CT) and positron emission tomography, tests that should improve on chest X-ray because of their much higher sensitivity *(20)*. Preliminary results raised doubts on efficacy and evidenced the risk of overdiagnosis for lung cancer screening *(21,22)*. Thus research is ongoing *(23,24)*, but it remains to be seen whether low-dose spiral CT screening also will gain diffusion despite uncertain benefit.

Time required to produce evidence is not only a problem confined to test proposed for previously unscreened cancer sites but also within existing screening interventions. Refinement of screening and development of new technologies are faster than trial planning and completion. For almost any existing screening, new technologies are proposed and often used in practice with variable supporting evidence. Minor screening variants, or "drift," borrowing the flu terminology, often consist of modification of existing tests: digital mammography is increasingly used instead of traditional film mammography in breast cancer screening *(25,26)*; digital mammography may be more sensitive in young women with dense breasts and allows telemammography, but it is more costly than traditional mammography. Further developments such as tomosynthesis are already on their way. The transition from guaiaco-FOBT to immuno-FOBT *(27)* or quantitative immuno-FOBT *(28)* for large bowel screening is another instance of the proposed introduction of a test variant to improve on a test with evidence of efficacy from randomized trials. Also, conventional pap smear test was replaced in some services with the more costly cervical cancer liquid-based cytology and automation-assisted cytology, and recently some comparative evidence was produced *(29,30)*.

Major screening variations, or shifts, include changing diagnostic test or performing additional, supplementary testing. For all the above-mentioned screenings, major variants are in general use (e.g., optical or virtual colonoscopy, and flexible sigmoidoscopy for large bowel screening) or find limited specific application (e.g., human papillomavirus [HPV] DNA testing in cervical cancer screening) *(31)* or are currently proposed (e.g., addition of magnetic resonance imaging [MRI] in breast cancer screening) *(32,33)*.

The addition of dermoscopy to the screening for skin melanoma *(34)* and, perhaps, of proteomics profiling to prostate cancer screening *(35)* represents additional examples for screening of uncertain benefit. It may seem that minor variants would require less evidence before implementation. If in general this holds true, we should remember that usually newer tests compare better (or are supposed to) than older tests in terms of sensitivity, specificity, or both.

If screening is judged mainly by its ability in reducing mortality and indeed it is (may be an example of tunnel vision *[36]*), sensitivity, the ability of the test to filter out persons with the disease or predisease from the target population, seems the most important feature of a test because it is related to the completeness of removal of target lesions and thus to the degree of mortality reduction achieved. Removal of potentially dangerous lesions also reduces the risk of litigation due to cancers missed at screening.

A higher sensitivity, however, may not be a positive feature if the main result of such improved sensibility is overdiagnosis. Indeed, in case of overdiagnosis, sometimes we are misclassifying pseudodisease as true disease, and thus we consider higher sensitivity (i.e., a desirable feature) what is in reality a lower specificity (i.e., a negative feature). In general, if we regard screening as filters the issue is more complex than merely changing sensitivity, we shall increasingly consider selectivity for progressive cancers as an important feature, and it is likely that this feature, if possible at all, will require the addition of new components to existing screenings *(37,38)*.

The case of prostate cancer and melanoma demonstrate that screening technologies may diffuse without definitive evidence of efficacy. This is blurring the distinction between proof of efficacy, deriving mostly from trials, and proof of effectiveness, deriving mainly from observational studies. Actually, controlled trials with observational research and screening evaluation systems provide relevant information on efficacy *(39,40)*.

Observational studies are increasingly based on data collection systems that are stable over time and include information on several variables related to person, disease, and health interventions. These evaluation systems derive information from screening provider or from cancer and mortality registries, and from other health archives. Thus, screening evaluation systems producing process and outcome indicators should play an increasing role in 1) providing evidence of efficacy, 2) providing timely evidence of efficacy for modified screening (evaluation of new technologies), and 3) monitoring effectiveness of local screening programs. The timely availability of detailed electronic information on screening histories, treatments, and health outcomes would ensure the healthy growth of screening as an efficient and effective tool in cancer control. Screening evaluation is important, however, because screening may produce harm, and because screening is a complex intervention that may be locally ineffective, despite

potential efficacy. Moreover evaluation is gaining importance because of the increasing role of secondary prevention in cancer control and the fast development and diffusion of innovative often costly diagnostic technologies.

5. Screening Types

The active search of cancer in asymptomatic individuals is pursued by two different models: the population-based or mass or organized screening model and the opportunistic, wild, or spontaneous screening model. The main distinction between organized and opportunistic screenings lies in the definition of a target population and personal invitation to participate *(41)*. However organized screenings differ from opportunistic screenings because of other features, including standardization of the screening, referral, and evaluation systems. Widely accepted criteria to define organized screenings were set forth by Hakama and colleagues in 1985 (**Table 6.1**; *(42)*).

Table 6.1
Defining criteria for organized screenings according to Hakama and colleagues[a]

a. The target population has been identified;
b. individual people are identifiable;
c. mechanisms are implemented to guarantee high coverage and attendance (e.g., a personal letter of invitation);
d. there are adequate field facilities for performing the screening tests;
e. there is a defined quality control program concerning how the tests are performed and interpreted;
f. adequate facilities exist for diagnosis and for the appropriate treatment of confirmed abnormalities;
g. there is a carefully designed and agreed upon referral system, an agreed link between the participant, the screening center, and the clinical facility for diagnosis of an abnormal screening test, for management of any abnormalities found, and for providing information about normal screening tests; and
h. evaluation and monitoring of the total program is organized in terms of incidence and mortality rates among those attending, among those not attending, at the level of the total target population. Quality control of the epidemiologic data should be established.

[a]Hakama M, Chamberlain J, Day NE, Miller AB, Prorok PC (1985). Evaluation of screening programmes for gynaecological cancer. Br J Cancer **52,** 669–673.

Although in the United States opportunistic screening is the prevailing paradigm for all screened sites *(43)*, in other western countries there seems to be a relationship between screening type and availability of evidence of efficacy: screening implemented after trials' results are mainly diffused as programmed interventions, and screening without evidence of efficacy are diffused as opportunistic. Thus, uncertainty on screening efficacy did not lead to clear-cut health policy decisions (i.e., refusal/acceptance), but often to the decision of not establishing organized screening programs. Spontaneous screenings, however, were able to diffuse in most western countries, although with variable intensity, due to general practitioners and specialists prescribing policies *(44)*. As Mandelsky et al. *(45)* note, "still, in many countries, health systems often directly or indirectly support such modes of delivery, for example, through fee-for-service reimbursement, which pay practitioners for the delivery of specific services." To be more precise, in the United States screenings are provided in several settings, such as private practice, health maintenance organizations, and academic medical centers, that may to varying extent met the criteria defining organized programs.

Cervical cancer screening is peculiar because it is deemed effective based on results from observational studies only: organized programs and opportunistic screenings are diffused, and often both screening types coexist within the same geographic area *(46,47)*. Evidence of efficacy from observational studies may be simpler to produce for screening aimed at detecting precancerous lesions than for screening detecting the target cancer at an earlier stage because incidence and mortality will vary in the same direction. For screening detecting invasive cancers, disentangling the effects of new treatments and early diagnosis on mortality trends adds a difficulty in the interpretation of observational studies.

Regardless, the coexistence of organized and opportunistic screening types makes available some comparison data for cervical cancer screening. Comparisons among countries with organized cervical cancer screening programs, countries with opportunistic screening activities, and countries with no screening, and before–after comparisons for geographic areas where programs were introduced following spontaneous screening, showed that opportunistic screening is effective compared with no screening, but it may be less effective in terms of lowering incidence of invasive cancer, adhesion to guidelines, and costs *(48–52)*. The lack of defined rules within the opportunistic screening context were associated to improper testing in terms of too frequent repetition of Pap test and inclusion of low-risk age groups (e.g., young women) and, as a consequence, increased risk of overdiagnosis and overtreatment and waste of resources *(53–55)*.

The introduction of organized screening allowed for better coverage because of active invitation. Women with lower access to health services because of socioeconomic, cultural, or other factors were shown to be less likely to receive screening. These subgroups may also be at higher risk of cervical cancer *(56)*. The undercoverage by cervical cancer screening among disadvantaged groups may be larger for opportunistic interventions.

Finally, process and outcome indicators produced within the organized screening programs should allow for timely individuation of problematic aspects and introduction of changes to improve effectiveness.

For cervical cancer, the adoption of the organized screening strategy was considered as a means to improve health and economics over spontaneous screening *(57–59)*. Thus, there is limited but concordant evidence that organized screenings showed some additional preventive benefit in the case of cervical cancer. Improved coverage after active invitation may be a major explanation for the better health outcomes achieved by organized programs *(60)*; additional explanations include the effect of quality control for screening components and of the evaluation systems established with organized programs.

Data from the United States, however, indicate that opportunistic screening in a different context may produce results comparable with organized screenings. In the United States, high coverage is reported for Pap tests by the Behavioral Risk Factor Surveillance System survey (http://apps.nccd.cdc.gov/brfss/display.asp?cat=WH&yr=2006&qkey=4426&state=US), with rates that are comparable with rates achieved by organized programs. This may be related to a diffuse positive attitude toward cancer screening *(61)*. A relevant decrease in incidence of cervical cancer also was reported since 1940s.

Reduced access to cervical cancer screening was indeed demonstrated in subpopulations with reduced access to care because of socioeconomic, cultural/ethnical, and demographic reasons, and targeted interventions were established to improve coverage in these subgroups *(61,62)*. Moreover, reduced cervical cancer screening due to the above-mentioned factors is not eradicated within organized programs, and it may periodically reemerge *(63,64)*.

Opportunistic screening was associated with improper testing and waste of resources *(45)*. Evidence on age limits and screening interval is poorly defined. There is still considerable debate on optimal age limits (particularly about the age when screening should start, with options ranging from 20 years or within 3 years after the start of sexual activity to 30 years of age) and interval between successive testing (with option ranging from annual to every 5 years). There is little reason because variability between organized programs should be superior to

variability within opportunistic activity. Moreover, incidence of cervical cancer shows high variability by birth cohort, particularly in young women due to variation in risk factors exposure; thus, effectiveness and efficiency of screening by age may vary over time accordingly *(65)*.

Literature comparing different screening strategies is sparse, and for cancer sites other than cervical cancer, almost lacking *(45)*. Lack of data depends on difficulties to evaluate opportunistic screening (individual screening histories are not available; thus, comparisons between screen detected and symptomatic cases are not possible), complexity in defining and comparing organizational models, and prevailing of a single screening type by cancer sites. Thereby, it is not possible to be conclusive on the superiority of one screening model over another model. Prostate cancer screening, however, in which annual testing is diffused even among oldest old, clearly shows that there are considerable risks of harm and waste of resources associated with uncontrolled opportunistic screening.

As concluding remarks, evaluation of the screening process and outcomes is important, and evaluation is simplified within organized screenings. Diffusion without evaluation and thus lack of production of new scientific knowledge may be considered a major limitation of the opportunistic model. Better control over expenses also may be achieved by organized programs. Rapid diffusion without initial investments and timely incorporation of new technologies may favor spontaneous screening but with some increased risk of introduction in clinical practice of ineffective and harmful interventions.

Flexibility of spontaneous screening may lead to intervention better tailored on individual needs but also to less than optimal care because of over- or undertreatment (i.e., management not following best evidence). The same quality of care is not ensured to all people receiving screening in the opportunistic context. Referral and evaluation systems, components of organized screenings, may reduce variability in quality of care offered after diagnosis and health outcomes. Conversely, the rigid design of the screening process within organized programs may have drawbacks. The choice of a rigid age limit that is usual within organized programs may exclude population subgroups from receiving an effective preventive intervention. This may be the case for breast cancer screening in women >69 years in most European countries *(66–68)*. Presently, the organized screening seems associated with some lack of adaptability: age limits for programs use arbitrary 5- or 10-year age classes for all organized screenings (such limits also may vary among different programs), and the testing strategy (e.g., screening interval) is the same for all the target population, but this one size choice does not necessarily lead to optimal fit for all ages and screened sub-

groups. For example, features of the breast gland in premenopausal women (i.e., higher density) may lower the sensitivity of mammography; an alternative testing strategy may include MRI or ultrasound for women with dense breasts or/and adoption of a shorter testing interval *(69)*. Carriers of genetic mutations leading to hereditary cancers also could also require tailoring of the screening strategy in term of choice of tests and periodicity of screening, taking into account the personal preferences of these high-risk persons. Other identifiable subgroups at higher or lower risk may benefit as well from specifically adapted screening strategies.

It is possible to consider under the heading of screening a third occurrence of diagnosis of cancer not deriving from investigation of cancer's symptoms, the unintended or side effect. Incidental diagnosis of cancer is not uncommon (**Table 6.2**). For example, investigation for a suspected prostate cancer may lead to diagnosis of asymptomatic urinary bladder lesions, and this is called diagnostic bias in studies investigating association between urinary and prostate cancers due to risk factors *(70)*. However, we should reserve the definition of unintended cancer screening to situations when systematic or frequent investigation are carried out, and this activity leads to a relevant change in cancer incidence, similar to change produced by the introduction of intended screenings. We can cite two instances of side effect screenings: the increase in prostate cancer incidence due to treatment of benign prostatic hypertrophy by transurethral resection (TURP) in the 1970s and 1980s *(71)* and the present increasing trend of thyroid cancer incidence.

In the United States, according to the estimates of Merrill et al. *(72)*, "the use of transurethral resection of the prostate (TURP) in recent decades to treat obstructive uropathy have made a substantial contribution to prostate cancer incidence rates and trends." Indeed, prostate cancer incidence was estimated to increase by 2% per year before the diffusion of PSA testing in the SEER area.

Incidental diagnosis of prostate cancer at TURP may be of some importance, but the use of this procedure is much lower than in the 1970s due to medical treatment of prostate hypertrophy

Table 6.2
Broad screening categories for distinguishing diagnostic activities in asymptomatic individuals (i.e., not investigated because of symptoms of the disease under screening)

1. Organized screenings
2. Opportunistic screenings
3. Involuntary or side effect screenings

and to preference for transrectal biopsies *(73,74)*. Increasing incidence of thyroid cancer since the 1970s was reported by cancer registries in many areas *(75,76)*. This trend is almost certainly due to changing medical practice, even if a concurrent true increase due to risk factors may not be ruled out *(77)*. Sensitivity of ultrasound techniques and use of fine-needle aspiration techniques together with histologic criteria for diagnosis were probably responsible for incidence increase *(78,79)*. Incidence increase in western countries mostly concerns papillary cancers and early stages, including microcarcinomas (i.e., <10 mm), a pattern very similar to the increase of superficial spreading and in situ skin melanoma associated with opportunistic screening diffusion.

6. Cancer Screening and Epidemiologic Indicators

Screening causes important modifications of epidemiologic indicators, particularly incidence and survival *(1)*.

6.1. Incidence

Screening activities influence incidence, at least temporarily. Indeed, first or prevalence screening usually determines a transient increase in incidence rates *(80)*. The identification of precancerous lesions or in situ cancers decreases incidence of invasive cancers to the extent that those lesions would undergo malignant transformation during the screened lifetime. Removal of premalignant lesions such as colonic polyps or cervical dysplasia does not produce immediate incidence decreases; for an effect on incidence rates to become apparent, the time between diagnosis at screening and the hypothetical time of diagnosis of invasive cancer in the absence of screening should pass for a sufficient quote of cases. Moreover, incidence variation depends on disease risk due to risk factors exposure and to the coverage achieved by the screening program *(81,82)*. Indeed, colorectal cancer screening trials reported a shrinking difference in the number of invasive cancers and a large difference in the number of adenomas *(83,84)*. If the screening test detects lesions that would remain undiagnosed in the absence of screening (overdiagnosis), incidence rates would increase and then remain stable over time (after adjusting for risk factors) at a level higher than before screening introduction.

6.2. Screening and Survival

Longer survival among screen-diagnosed cases is intrinsic to the screening concept. A screening test must have the ability to detect the disease earlier in its natural history; thus, the time corresponding to the anticipation of diagnosis is added to survival time for

each screen-detected case. Although the individual anticipation is not determinable because we do not know when symptomatic diagnosis would have occurred in absence of screening, methods to estimate the average lead time were proposed *(85–90)*. For example, lead time estimates for breast cancer range from 1 to 4 years and for prostate cancer from 5 to over 10 years; lead times estimates vary also by age class. In observational studies, the average lead time also may vary by population due to differences in awareness and access to diagnostic facilities. Survival rates for screened cases tend to be higher also because of length time bias, that is a selective feature of screening strategies leading to a difference distribution of progression time between screen-detected and nonscreen-detected cases. Length time bias refers to the lower probability to detect aggressive lesions with short sojourn times at screening; the choice of the interval between successive testing should influence the ability to detect such aggressive lesions. On the contrary, slowly progressive lesions are preferentially detected at screening; the existence of a pool of very slowly or nonprogressive lesions that would remain undiagnosed in the absence of screening exists (cases overdiagnosed at screening) causes a fake increase in survival rates. All overdiagnosed cases result cured of the disease.

Nonrandomized studies are also likely to suffer of prognostic selection bias *(91)*. People participating in screening are often healthier, more affluent, and of higher cultural level on average than nonparticipating people and they tend to have better health outcomes. Thereby, length bias and self-selection bias contribute to prognostic bias favoring screen-detected cases, i.e., an apparent survival advantage for cases diagnosed at screening. Because of negative prognostic selection, survival rates may not be simply interpreted even for nonscreen-detected cases. The relevance of selection bias may vary across populations depending on interventions targeted at increasing participation among population subgroups and overall adhesion to screening (should all target population participate in screening there would be no prognostic bias).

There is room also for a different behavior of caring health system favoring screen-detected cases *(92)*. Due to referral systems and treatment pathways designed for screened cases within organized screening programs, time to receive diagnostic confirmation and treatments could be lower for screen detected cases and also care received may be more adherent to current guidelines because of evaluation. The opportunity for more timely and effective care among screened could be named “care bias.” This may apply to some particular programs and not to others.

In general a screening intervention not associated with increasing survival is ineffective but the contrary does not necessarily hold true. Prolonging survival is a necessary but not sufficient feature of an effective screening test. Screenings leading

mostly to diagnosis of premalignant lesions, such as screening to prevent cervical cancer, represent an exception because they have a major influence on incidence while survival rates change little.

Colorectal cancer screening should influence survival first due to early diagnosis of invasive cancers and incidence later because of removal of colonic dysplastic adenomas destined to undergo malignant transformation in the absence of screening. Survival rates often remain higher among screened than among control groups even after adjusting for stage and other variables *(93–96)*.

6.3. Mortality

Mortality reduction is the ultimate goal of screening and it is the main indicator to evaluate screening efficacy. Problems in using mortality for screening evaluation purposes are in part different between the trial and observational study context. The arbitrary assignment of cause of death may be responsible for biased trial results. All-cause mortality is not affected by cause of death assignment; thus, it is proposed as a more robust indicator of screening efficacy *(97)*. However, the disease under study usually determines a small fraction of deaths; thus, all-cause mortality seems not a sensitive endpoint in cancer screening *(83)*.

In observational studies, one of the main limitations in using mortality to assess screening effectiveness, lies in the influence of cases diagnosed years before on the mortality rate so that mortality trends after the introduction of a screening also will depend on cases incident before the intervention *(98)*. The incidence period relevant to determine the mortality rate in a certain year varies by cancer site (i.e., it is shorter for rapidly lethal cancers). Calculation of mortality based on cases incident over selected periods (e.g., refined mortality) was proposed to account for this problem, but this method requires additional individual information *(99)*. Misclassification of cause of death is also of concern in observational studies *(100,101)*. Lack of resolution also exists for mortality data, and the attribution of a death to cervix or corpus uteri is an example of such problem *(102)*. Finally, mortality with its evidence, strength, and availability may be exceedingly considered in the assessment of cancer screening and obscure the importance of other endpoints, usually more difficult to be measured, like disability, quality of life, harm.

7. Screening, Overdiagnosis, and Overtreatment

Overdiagnosis is commonly defined as the detection of cases that would have not progressed to symptomatic diagnosis in the absence of screening. Paci and Duffy stated that “overdiagnosis

is largely an epidemiological concept, because there is no marker today to classify a cancer as a pseudodisease" *(103)*. However, the wide definition of overdiagnosis reported above is not congruent with the term "pseudodisease" and the possibility of distinguishing pseudocancers using a marker. Some progressive cancers remain undiagnosed because a cause of death other than the cancer occurs during the asymptomatic phase *(1)*. Both the duration of the asymptomatic phase and competitive causes of death influence overdiagnosis *(103)*.

A screening can detect progressive lesions (as such not different from other "true" cancers) that would otherwise go undiagnosed because of competitive causes of death; thus, some overdiagnosis is inherent in the screening process. Overdiagnosis is a quantitative concept. If a population with a high prevalence of latent lesions with sojourn time longer than life expectancy undergoes screening, then overdiagnosis results that is a relevant and lasting unbalance between incidence before and after screening diffusion. Also, the mortality/incidence ratio is typically decreased. Among the oldest old, the role of competitive causes of death prevails, and among young adults the nonprogressive feature of some cancers plays a major role as determinant of overdiagnosis. In the elderly, the scope for overdiagnosis is also higher because, for many epithelial cancers, the prevalence of cancer in the asymptomatic phase is higher. Cancers with a stay in the preclinical phase less than life expectancy will have a lead time if detected at screening, and cancers with a stay longer than life expectancy will be overdiagnosed. Because overdiagnosis by definition does not produce health benefit but harm and waste of resources, it is important to identify and quantify the problem by age and cancer site, and to reduce it as much as possible. Underlying incidence trend and lead time make it difficult to identify overdiagnosis, and they are responsible for highly variable estimates of the level of overdiagnosis.

There are two ways to reduce overdiagnosis: exclusion of age classes (organized screenings) or individuals (opportunistic screenings) with limited life expectancies and identification of "pseudodiseases" not requiring treatment. A deep knowledge of the natural history of screened cancers is necessary to distinguish benign and nonprogressive lesions (with progression time much exceeding the individual life expectancy). Indeed, it would be hard both for the health professionals and for affected persons to leave a cancer untreated unless its nonprogressive nature would be almost certainly demonstrated. Because the term pseudodisease entails a biologic difference between true cancers (those progressing to symptomatic phase) and nonprogressive lesions, it is more suited for misclassified benign lesions, as may be the case for a quote of skin melanomas, and for nonprogressive cancers than for all overdiagnosed cancers.

Little doubt exists that prostate cancer is presently overdiagnosed. Autopsy studies found a very high prevalence of latent prostate cancers before the screening diffusion (about 30% in men >50 years) *(104–107)*. Latent cancers also increase with age, and >60% of men ~70 years harbor a prostate cancer *(108)*. Incidence of prostate cancer increased dramatically in many countries as a consequence of PSA opportunistic screening in the 1990s, whereas incidence was already increasing as a side effect of TURP diffusion. Despite the risk of overdiagnosis and uncertain benefit, PSA screening is widely diffused (e.g., >50% of men >40 years old in the United States according to the 2006 Behavioral Risk Factor Surveillance System [BRFSS] survey; more than half Canadian men >50 years *[109]*). PSA based screening is more popular than the effective large bowel FOBT-based screening *(110,111)*. PSA screening is not recommended in men with life expectancy <10 years *(112,113)*. Both the huge prevalence of latent cancers and the limited life expectancy strongly discourage testing in older men, but PSA testing also is frequent in this age group *(114–116)*. Estimated lead times ranging from 5 to >10 years suggest intervals between successive testing longer than annual *(88)*, but annual testing is frequent *(110,117)*; this is likely to increase the risk of overdiagnosis.

Screening modifications to detect more cancers (e.g., lowering PSA threshold with respect to actual 4-ng/ml value) are frequently advocated or simply adopted, not recognizing the influence of these choices on overdiagnosis *(118–122)*. Increasing the number of random biopsies after "abnormal" PSA results until saturation, and other strategies to increase sensitivity are actively developed and proposed as well *(123–126)*.

Estimates of overdiagnosis for prostate cancers vary, but they are almost always larger than 20%; values about >50% are commonly reported; moreover, in men >75 years, the estimates are substantially higher *(88–90,127,128)*.

A corollary of such levels of overdiagnosis is that test features' such as sensitivity are difficult to interpret and may be misleading. Treatment for prostate cancer including radical prostatectomy and radiotherapy has important side effects. Perioperatory mortality is rare (about 1/1000) and has decreased since screening diffusion, but other disabling conditions (e.g., severe urinary incontinence and sexual impotence) remain frequent *(129–131)*.

Overdiagnosis is relevant also to other screenings. Yet among screenings of uncertain benefit, opportunistic search for melanoma of the skin is likely to determine overdiagnosis as judged by inci dence trends *(132,133)* (*see* **Table 6.1**). In this case, the lack of firm histologic criteria for identifying malignant lesions contributes to the phenomenon *(134)*. The contribution of pseudodisease to the incidence of melanoma has not yet been estimated to our knowledge. Overdiagnosis may be less of concern when treat-

ment has not important adverse effects as for skin melanoma, and for premalignant lesions like cervical dysplasia or intestinal polyps. Indeed, we accept the overdiagnosis of intestinal polyps simply as a consequence of screening, but we question that of prostate cancer, and the difference between the two cases lies in the consequences of treatment and the labeling as a cancer patient. Fear of litigation in some screening contexts represents an additional factor that may lead to aggressive investigation of any not completely reassuring diagnostic test result and possibly of overstating suspicious and uncertain findings *(132,135)*.

The terms overdiagnosis and overtreatment were recently referred to thyroid cancer in a German paper, although they are usually reserved for the screening context *(136)*. There are similarities between thyroid cancer and, for example, prostate cancer, regarding the existence of a prevalence pool of cases demonstrated by autopsy studies *(137–139)*, the circumstance that incidence at some time point showed an abrupt change that cannot be explained by risk factors, and that cancer detection often leads to radical surgical removal of the gland and to radiotherapy treatment. Markers to filter progressive lesions from likely overdiagnosed cases and "pseudocancers" are lacking for both prostate and thyroid cancer. However, there are also differences because mortality from thyroid cancer is much lower than for prostate cancer. In addition, the recent increase in thyroid cancer incidence was related to the diffusion of new diagnostic techniques (i.e., they are with an ugly word called "incidentalomas") and not to active search for cancer *(140,141)*. Incidental microcarcinomas and papillary cancers represent a clinical problem and efforts should be made to identify features characterizing clinically relevant disease and to minimize the risk of overtreatment *(142)*.

Overdiagnosis is also of concern for breast cancer screening (*see* **Subheading 8.**). Estimates of the level of overdiagnosis associated with mammography screening vary from <5% *(143)* to 50% *(144)*. A separate issue is represented by ductal carcinoma in situ (DCIS), a pathological entity diagnosed almost only at screening, where uncertainties regarding its natural history and treatment, imply a risk of overtreatment *(145–147)*.

Potential for overdiagnosis exists for screening proposed or under study, such as lung cancer screening. Overdiagnosis and overtreatment may explain some of the findings of the lung cancer screening trial based on chest X-ray in which incidence and survival turned out to be markedly increased in the screened group that also showed nonsignificantly higher mortality *(21,148)*. Evidence of the existence of a latent prevalence pool from autopsy studies is available also for lung cancer *(149)*. Available data on the low-dose spiral CT scan screening actually under study as lung cancer screening test showed much higher sensitivity than

chest X-ray, raising justifiable concern for the risk of overdiagnosis *(150,151)*.

Thus, overdiagnosis is overall a major problem in cancer screening. It is central for prostate cancer because of both the level of overdiagnosis and the consequences of overtreatment. It may be a relevant problem also in breast cancer, but in this case, more research is needed to reduce variability in estimates. As to melanoma, non-negligible overdiagnosis exists, and it should be quantified even if overtreatment has less severe consequences than for prostate or breast cancer.

8. Short Description of Screening Features, Diffusion, Achievements, and Limitations for Skin Melanoma and Breast Cancer

8.1. Breast Cancer

8.1.1. Background

Breast cancer is the most frequent cancer among females: it is estimated to account for about 22% of all female cancer cases wordwide. Incidence is high particularly in western countries *(152)*. Breast cancer is also the leading cause of cancer deaths among women in most western countries (in the United States and Canada lung cancer ranks first) (*[153,154]*; Canadian Cancer Society/National Cancer Institute of Canada: Canadian Cancer Statistics 2007, Toronto, Canada, 2007; http://www.cancer.ca/vgn/images/portal/cit_86751114/36/15/1816216925cw_2007stats_en.pdf).

Although primary prevention may gain importance in the future *(155–157)*, presently screening and treatment are the main interventions to control breast cancer. Mortality from breast cancer is declining in most western countries; for example, a 2% annual reduction was reported for the European Union starting in 1995 *(158)*; and in the Untied States, the decline started earlier, in 1990; peaked in the period 1995–1999; and it was still decreasing until 2003, by 1.4% per year *(154)*.

Both advances in treatment and screening diffusion are likely to have contributed to reduce breast cancer mortality in many western countries, but the respective merits are controversial.

8.1.2. Screening Type

Outside the United States, mammography screening is provided mostly within the context of organized screening programs *(67,68,159,160)*. Most organized screenings adopted age limits, and the most used is 50–69 years. In the United States, mammography screening is diffused also among older women *(161)*.

8.1.3. Evidence

All but one of seven trials (the Canadian trial is indeed two trials studying different age groups) showed some mortality reduction associated with mammography screening. Breast cancer mortality reduction was significant in only one trial *(162–168)*.

Two systematic reviews, respectively, from U.S. Preventive Task Force (USPTF) and Cochrane collaboration found important methodological limitations in all of the trials. The USPTF estimated the relative breast cancer mortality reduction to be 16% (95% confidence interval [CI] from 23 to 9%) and the Cochrane review 20% (95% CI from 30 to 9%) *(12,13)*. In the recent Cochrane evidence update, doubts are cast on the overall harm/benefit ratio of breast cancer screening.

Many observational studies tried to relate mortality trends with mammography screening. Besides underlying variations in disease risk *(169)* and general problems of quality of mortality data *(170)*, treatment improvements and unassessed opportunistic screening diffusion hindered a direct interpretation of trends. Within Europe, screening diffusion does not entirely explain mortality time trends *(98,171,172)*. Little change in overall mortality was observed after the introduction of organized screening programs in countries such as Sweden and Finland. A possible explanation for observed trends, taking into account also incidence and survival trends, is that residual health gain that organized screening may achieve are likely to depend on previous diffusion of opportunistic activities, level of population awareness and accessibility to health services, and quality of care *(98)*. Differences in study design and analysis are also responsible for large variability in estimates (applying various methods to Danish data resulted in estimates ranging from 25% reduction to 6% mortality increase) *(99)*. A recent mode-lling effort tried to disentangle the respective contribution of treatment and screening on breast cancer mortality reduction in the United States over the period 1975–2000 and attributed to screening a median of 15% for breast cancer mortality reduction, with estimates from different models ranging from 7.5 to 22.7% *(173)*.

8.1.4. High Risk and Screening Evolution

The increasing use of digital mammography in service screening and of computer-aided detection methods should not lead to major changes. Other diagnostic technologies, including ultrasound (US) and MRI, may gain a role in screening among premenopausal women at high risk of breast cancer because of family history, breast cancer susceptibility gene mutation (BRCA1 and BRCA2), dense breasts, and recent diagnosis of controlateral breast cancer *(174,175)*.

Parenchimal breast density is a mammographic feature *(176)* associated with two- to six-fold increased breast cancer risk *(177)* and with higher risk of false-negative mammography results *(178)*; digital mammography and annual testing were proposed to improve screening sensitivity in high-risk premenopausal women *(70,179)*. The prevalence of dense and even extremely dense breasts was found high in young premenopausal women

(180) and was much higher than the prevalence of BRCA genetic susceptibility (<5%).

Prove of effectiveness in reducing mortality and even of diagnosis at an earlier stage is lacking for US and MRI *(181,182)*. Thus, the use of these technologies rests on their capability to improve on mammography sensitivity and on the assumption that the probability of cure will be improved for these cancers as for the other screen-detected cases *(183)*. MRI could be responsible for less false-positive results than US *(184,185)*. With respect to annual mammography screening, MRI is more sensitive but also more costly and produces more false-positive results in high-risk populations such as BRCA mutation carriers *(186,187)*. Recently, the American Cancer Society recommended the use among women with 20–25% or greater lifetime risk of breast cancer, including women with strong family history but not among women with extremely dense breasts at mammography *(188)*.

8.1.5. Risk of Harm

Actually, it is not possible to distinguish between progressive cancer and pseudodisease at the individual level; should it be possible, overdiagnosis would represent much less of a problem. Thus, the very presence and quantity of overdiagnosis must be estimated, and various estimates are produced depending on the adopted methodology and assumptions. Overdiagnosis is consistent with survival advantage for screen-detected cases after adjusting for stage and grading and with no incidence decrease in age groups not participating in screening *(93,144)*. Estimates of overdiagnosis associated with mammography screening were produced from trials and service screening data. Presently, there is a general agreement that mammography screening is associated with overdiagnosis, but estimates of this phenomenon range from as little as <5% *(189)* to as much as 60% *(144)*.

Estimates of overdiagnosis from service screening also may reflect true differences in prevalence of slowly progressive lesions (i.e., overdiagnosis may be more frequent in high incidence Nordic countries than in Italy) *(143,190)*.

Diffusion of mammography leads to a large increase in diagnoses of DCIS of the breast *(145)*; both potential for progression and treatment are uncertain for these lesions. If a treatment producing no better health outcomes than no treatment is considered overtreatment, there should be scope for overtreatment of DCIS *(191)*.

The importance of overdiagnosis for breast cancer must not be underestimated, because although the disease is almost as accessible as melanoma, it is much more frequent, and commonly used treatments have important side effects. For example, radiotherapy was shown to increase the risk of subsequent lung cancer and hearth diseases *(192,193)*. Axillary dissection causing lymphoedema is associated

with considerable disability *(194)*; the incidence of lymphoedema is lower for the sentinel lymph node dissection technique.

8.1.6. Role of Screening

Breast cancer screening contributed to mortality decline observed in many western countries. However, the contribution of screening to cancer control is likely to vary by age and health system. Data are insufficient to compare opportunistic and organized breast cancer screening. Among European Nordic countries, the introduction of organized screening seems to have made little difference *(195,196)*. In the United States, high mammography rates were achieved by opportunistic screening, but participation was lower among uninsured and disadvantaged women *(197)*. Comparison between United Kingdom and some U.S. programs showed higher recall rates in the United States, with similar detection rate *(198)*. In Europe and other western countries, however, women older than 70–75 years did not receive screening because of age limits adopted within organized screenings. Elderly women have a high prevalence of slowly growing tumors, but breast cancers are often diagnosed at an advanced stage, probably because of reduced access to health services. Elderly women are often undertreated and thus receive little benefit from recent advances in treatment *(199)*. Breast cancer screening for elderly women is considered a cost-effective intervention, and its diffusion would contribute to mortality reduction *(169,200,201)*.

As for other screening, research to produce less variable estimates of overdiagnosis and to distinguish progressive lesions from pseudodisease is urgent. The definition of proper treatment for DCIS is of concern, and it could contribute to avoid overtreatment. The addition of MRI to mammography for premenopausal women with extremely dense breasts could produce improvements both in terms of health outcomes and reduction of interval cancer rate in young women (also a cause of litigation), but it is costly and solid evidence is not yet available.

8.2. Skin Melanoma

8.2.1. Background

Melanoma of the skin is unhomogeneously distributed worldwide: the highest incidence rates are estimated among fair-skinned White populations in Australia and New Zealand, the United States, and northern Europe *(152)*. The interpretation of incidence rates and trends for melanoma is not straightforward because risk factors may have led to increasing incidence, and this, in turn, to increased population awareness and opportunistic screening activities. Moreover, difficulties in establishing the histologic diagnosis of malignant melanoma do exist *(134)* and there is variation in completeness of registration by cancer registries *(202)*. Due to early diagnosis and risk factor exposure, skin melanoma ranks among the most frequent cancer sites in Australia (fourth in males and third in females; Australian Institute of Health and Welfare [AIHW] & Australasian Association of Cancer

Registries 2004. Cancer in Australia 2001. AIHW cat. no. CAN 23. Canberra: AIHW [Cancer Series no. 28]; http://www.aihw.gov.au/publications/can/ca01/ca01.pdf) and the United States (sixth in males and fifth in females *[154]*). Mortality rates show much less geographical variability *(152)*. Even in high incidence countries, melanoma is not among the most frequent causes of cancer deaths. Diverging trends over time were observed: with increasing rates in European countries, except Scandinavia; stable rates in the United States; and stable and recently decreasing rates in Australia and northern Europe *(203,204)*.

8.2.2. Screening Type

Screening for melanoma by skin examination is diffused as an opportunistic screening. Peculiar of this screening is the relation between awareness and diagnostic test; indeed, the same skin examination may be self-performed or received by a dermatologist. Moreover, simple information to suspect malignancy (e.g., the mnemonic ABCDE acronym standing for asymmetry, border irregularity, color variegation, diameter ≥ 6 mm, elevation or evolution) is often given together with information on screening.

8.2.3. Evidence

No evidence from randomized controlled trials is available for melanoma secondary prevention *(205)*. A trial to evaluate screening based on whole skin examination is ongoing in Australia, but it will take years for results to be available *(10)*. Evidence of efficacy from observational studies is also controversial, despite large diffusion of screening activities in some areas over the past few decades (e.g., the American Academy of Dermatology is being offering free skin examination since 1985; in Queensland, Australia, early detection efforts date back to the early 1960s) *(206)*. There is evidence that early diagnosis increases incidence of thin, early stage lesions, and in situ melanomas but not that it decreases incidence rate of thick melanomas.

A much similar pattern with stable mortality rates in U.S. and Australian states, stable incidence of thick melanomas, and increased overall incidence due to in situ and thin melanomas was interpreted as evidence for early detection by Coory et al. *(206)* (early diagnosis effect counterbalance the influence on mortality of a real incidence increase) and against early detection by Welch et al. *(119)*.

Moreover, some mortality decrease was reported in Australian and Scandinavian mortality rates, and it was attributed to early diagnosis rather than to primary prevention because incidence was not decreasing *(204,207)*. It is to note, however, that wild screening may have masked an incidence reduction *(203)*, that one should look at incidence rates of thick melanomas instead of overall incidence, and it remains to be explained why it took so long for secondary prevention to produce its effect *(204)*.

Thus, it is not presently clear whether observed melanoma mortality trends may be better explained with sun exposure intervention, early diagnosis, or a combination.

8.2.4. High Risk

Most studies agree that skin surveillance is appropriate for high-risk people *(205)*. Total skin photographs and perhaps dermoscopy (see Subheading 8.2.5) may be added to periodic examination for these individuals. Persons with atypical mole syndrome, multiple multiple melanocytic naevi,, familial cutaneous malignant melanoma, or a history of cutaneous malignant melanoma are considered at high risk *(208)*.

8.2.5. Screening Evolution

Dermoscopy or dermatoscopy has been proposed to improve sensibility and specificity of skin examination *(209)*. The use of dermoscopy may reduce the benign/malignant ratio of excisions for suspected melanoma *(210)*. Concern was expressed as to the possibility that this technique may further inflate incidence of melanoma *(211)*.

8.2.6. Harm of Screening

Skin examination leads to the excision of benign lesions due to suspect of melanoma. The lack of firmly established criteria for histologic diagnosis of melanoma is likely to lead to misclassification of some benign lesions as malignant; this phenomenon may be increased in the context of screening activities due to pathologists' fear of for the legal consequences of a missed melanoma *(205)*. Lower mortality/incidence ratios are reported for high-incidence countries where skin melanoma ranks among most frequent cancer sites. Overdiagnosis due to misclassification and diagnosis of nonprogressive lesions is quite likely for skin melanoma *(119)*. Because the treatment of skin melanoma relies essentially on surgical excision, removal of benign lesions and overdiagnosis of melanoma seem not to cause major concern. Adverse effects of screening include scarring and fear. Consequences of overdiagnosis include being labeled as a melanoma patient and classified into a high-risk group. Indirect effects of early diagnosis of melanoma may include waste of resources and diverting attention from effective primary prevention strategies, should it prove ineffective.

8.2.7. Role in Melanoma Control

Regular skin examination may be useful among selected high risk persons *(205)*. Reduced sun exposure (particularly avoidance of sunburns in childhood) and early diagnosis may have influenced recent mortality trends in some countries and thus may play a role in melanoma control. However, the respective merits of primary and secondary prevention, if any, are yet controversial *(212,213)*.

Opportunistic screening and early diagnosis in general are almost sure to be responsible for non-negligible overdiagnosis

and thus overtreatment. Research to define firm criteria in the diagnosis of melanoma and to distinguish between progressive and nonprogressive cancers is urgent.

Most of the increase in mortality rates from melanoma observed in western countries is explained by increasing mortality in the elderly; an unfavorable trend continues in this age group also in countries reporting recent decreases in mortality. Strategies to control melanoma also may focus on the elderly population.

9. Conclusions: Screening Role in Cancer Control and Future Perspectives

Screening has an important and increasing role as a tool for cancer control. Mortality is declining in many western countries for cancer sites interested by screening activities. The contribution of ongoing screening to cancer control is much more controversial.

Systematic reviews estimated mortality reduction associated with breast and large bowel cancer to be about 15%. Mortality reduction that may result from prostate cancer screening, if any, is unlikely to exceed 20% *(214,215)*. Thus, actual screenings are providing a valuable contribute to decrease cancer mortality, but they do not seem able to achieve a major reduction of cancer burden. Several strategies are actively pursued to obtain mortality reductions as high as possible with ongoing screening interventions, including active development of new diagnostic tests, improving adhesion both in general population and among population subgroups (i.e., people with reduced access to screening). Moreover, screening is proposed for a number of additional cancers sites.

Screening, however, is also responsible for complex issue and important adverse effects. Here, overdiagnosis and evaluation are mentioned as two major problems that earned to actual screening the definition of double-edged sword *(61)*. The observed mortality decline after the introduction of a screening often is neither as fast as the incidence raise nor of a comparable entity. Overdiagnosis is an important adverse effect for many screenings, but it is under-recognized for several reasons. The main obstacle in the recognition of overdiagnosis is that actually it exists only in cancer figures (i.e., overdiagnosis is a guess based on strange behaviors of the cancer indicators incidence and survival); clinicians and screened individuals only perceive cancers. Indeed, presently it is not possible to distinguish between overdiagnosed and true cancers; even worse, this distinction is simply not possible for a quote

of cancers "overdiagnosed," mainly because of competitive causes of death that are unlikely to be biologically different from progressive cancers. Although two strategies may be considered to reduce overdiagnosis (i.e., limiting access to screening on the basis of a comparison between average lead time and life expectancy and developing new markers to identify pseudodisease), screening evolution often goes in the opposite direction by increasing the sensitivity of screening tests and adhesion among elderly people with limited life expectancies. This problem prevails in the unorganized opportunistic context, and it is of particular concern for prostate cancer. Overdiagnosis is of concern when causes overtreatment. Treatment of overdiagnosed cases does not produce health benefit but only harm, so that it increases the harm/benefit ratio and causes considerable waste of resources.

Quickly evolving diagnostic technology is a positive feature of screening, as is the attitude toward screening of health professionals and most of the general population. However, these two seemingly unrelated circumstances may lead to uncontrolled raise in healthcare costs if trials and observational studies do not provide information of efficacy of interventions. The idea that early detection is effective leads to the proposal of screening for many new cancer sites (even to the application of diagnostic technologies such as CT to find cancer at any site) and to the diffusion of screenings of uncertain benefit and costly techniques of unproven superiority. Because the evaluation of new or modified interventions requires years, it is no more an activity preceding the widespread diffusion of screening activities (evaluation of efficacy in well designed clinical trials) and then monitoring field results, but it parallels screening evolution in the two forms of experimental and observational studies. Evaluation, however, is a crucial activity that should be enforced to produce results of high quality and as timely as possible because evaluation allows for positive selection of effective interventions, avoiding uncontrolled and increasing waste of resources and stable diffusion of ineffective, harmful interventions.

Health policies aimed at maintaining and improving the role of screening as a cancer control tool should promote the following:

- definition and dissemination of guidelines to ensure that screening and subsequent treatment are implemented following best available evidence;
- research on natural history of cancer *(216)* and biomarkers to distinguish progressive cancers from pseudodisease; and
- screening evaluation both in terms of randomized trials and observational studies. Evaluation based on observational studies would take great advantage from establishing data sources containing detailed and readily available information on cancer diagnosis and care.

References

1. Kaposi, M. (1872) Idiopathisches multiples Pigmentsarkom Der Haut. *Arch Dermatol Syph* 3, 265–273.
2. Iscovich, J., Boffetta, P., Franceschi, S., et al. (2000) Classic Kaposi sarcoma: epidemiology and risk factors. *Cancer* 88, 500–517.
3. Cohen, A., Wolf, D. G., Guttman-Yassky, E., et al. (2005) Kaposi's sarcoma-associated herpesvirus: clinical, diagnostic, and epidemiological aspects. *Crit Rev Cl Lab Sci* 42, 101–153.
4. Dourmishev, L. A., Dourmishev, A. L., Palmeri, D., et al. (2003) Molecular genetics of Kaposi's sarcoma-associated herpesvirus (human herpesvirus-8) epidemiology and pathogenesis. *Microbiol Mol Biol Rev* 67, 175–212.
5. Cook-Mozaffari, P., Newton, R., Beral, V., et al. (1998) The geographical distribution of Kaposi's sarcoma and of lymphomas in Africa before the AIDS epidemic. *Br J Cancer* 78, 1521–1528.
6. Sitas, F., Newton, R. (2001) Kaposi's sarcoma in South Africa. *J Natl Cancer Inst Monogr* 1–4.
7. Sissolak, G., Mayaud, P. (2005) AIDS-related Kaposi's sarcoma: epidemiological, diagnostic, treatment and control aspects in sub-Saharan Africa. *Trop Med Int Health* 10, 981–992.
8. Zmonarski, S. C., Boratynska, M., Puziewicz-Zmonarska, A., et al. (2005) Kaposi's sarcoma in renal transplant recipients. *Ann Transplant* 10, 59–65.
9. Marcelin, A. G., Calvez, V., Dussaix, E. (2007) KSHV after an organ transplant: should we screen? *Curr Top Microbiol Immunol* 312, 245–262.
10. Goedert, J. J. (2000) The epidemiology of acquired immunodeficiency syndrome malignancies. *Semin Oncol* 27, 390–401.
11. Cheung, T. W. (2004) AIDS-related cancer in the era of highly active antiretroviral therapy (HAART): a model of the interplay of the immune system, virus, and cancer. "On the offensive – the Trojan Horse is being destroyed" – Part A: Kaposi's sarcoma. *Cancer Invest* 22, 774–786.
12. Bower, M., Palmieri, C., Dhillon, T. (2006) AIDS-related malignancies: changing epidemiology and the impact of highly active antiretroviral therapy. *Curr Opin Infect Dis* 19, 14–19.
13. Cheung, M. C., Pantanowitz, L., Dezube, B. J. (2005) AIDS-related malignancies: emerging challenges in the era of highly active antiretroviral therapy. *Oncologist* 10, 412–426.
14. Beral, V., Peterman, T. A., Berkelman, R. L., et al. (1990) Kaposi's sarcoma among persons with AIDS: a sexually transmitted infection? *Lancet* 335, 123–128.
15. Archibald, C. P., Schechter, M. T., Le, T. N., et al. (1992) Evidence for a sexually transmitted cofactor for AIDS-related Kaposi's sarcoma in a cohort of homosexual men. *Epidemiology* 3, 203–209.
16. Vogel, J., Hinrichs, S. H., Reynolds, R. K., et al. (1988) The HIV tat gene induces dermal lesions resembling Kaposi's sarcoma in transgenic mice. *Nature* 335, 606–611.
17. Ensoli, B., Salahuddin, S. Z., Gallo, R. C. (1989) AIDS-associated Kaposi's sarcoma: a molecular model for its pathogenesis. *Cancer Cells* 1, 93–96.
18. Chang, Y., Cesarman, E., Pessin, M. S., et al. (1994) Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science* 266, 1865–1869.
19. Russo, J. J., Bohenzky, R. A., Chien, M. C., et al. (1996) Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus (HHV8). *Proc Natl Acad Sci USA* 93, 14862–14867.
20. Moore, P. S., Chang, Y. (1995) Detection of herpesvirus-like DNA sequences in Kaposi's sarcoma in patients with and without HIV infection. *N Engl J Med* 332, 1181–1185.
21. Cesarman, E., Chang, Y., Moore, P. S., et al. (1995) Kaposi's sarcoma-associated herpesvirus-like DNA sequences in AIDS-related body-cavity-based lymphomas. *N Engl J Med* 332, 1186–1191.
22. Soulier, J., Grollet, L., Oksenhendler, E., et al. (1995) Kaposi's sarcoma-associated herpesvirus-like DNA sequences in multicentric Castleman's disease. *Blood* 86, 1276–1280.
23. Codish, S., Abu-Shakra, M., Ariad, S., et al. (2000) Manifestations of three HHV-8-related diseases in an HIV-negative patient: immunoblastic variant multicentric Castleman's disease, primary effusion lymphoma, and Kaposi's sarcoma. *Am J Hematol* 65, 310–314.
24. Ascoli, V., Signoretti, S., Onetti-Muda, A., et al. (2001) Primary effusion lymphoma in HIV-infected patients with multicentric Castleman's disease. *J Pathol* 193, 200–209.
25. Sukpanichnant, S., Sivayathorn, A., Visudhiphan, S., et al. (2002) Multicentric Castleman's disease, non-Hodgkin's lymphoma, and Kaposi's sarcoma: a rare simultaneous occurrence. *Asian Pac J Allergy Immunol* 20, 127–133.
26. Kasolo, F. C., Mpabalwani, E., Gompels, U. A. (1997) Infection with AIDS-related herpesviruses in human immunodeficiency virus-negative infants and endemic childhood

Kaposi's sarcoma in Africa. *J Gen Virol* 78(Pt 4), 847–855.

27. Sarmati, L. (2004) HHV-8 infection in African children. *Herpes* 11, 50–53.
28. Andreoni, M., El-Sawaf, G., Rezza, G., et al. (1999) High seroprevalence of antibodies to human herpesvirus-8 in Egyptian children: evidence of nonsexual transmission. *J Natl Cancer Inst* 91, 465–469.
29. Wang, Q. J., Jenkins, F. J., Jacobson, L. P., et al. (2001) Primary human herpesvirus 8 infection generates a broadly specific CD8(+) T-cell response to viral lytic cycle proteins. *Blood* 97, 2366–2373.
30. Luppi, M., Barozzi, P., Schulz, T. F., et al. (2000) Bone marrow failure associated with human herpesvirus 8 infection after transplantation. *N Engl J Med* 343, 1378–1385.
31. Sarid, R., Pizov, G., Rubinger, D., et al. (2001) Detection of human herpesvirus-8 DNA in kidney allografts prior to the development of Kaposi's sarcoma. *Clin Infect Dis* 32, 1502–1505.
32. Marcelin, A. G., Roque-Afonso, A. M., Hurtova, M., et al. (2004) Fatal disseminated Kaposi's sarcoma following human herpesvirus 8 primary infections in liver-transplant recipients. *Liver Transpl* 10, 295–300.
33. Renne, R., Zhong, W., Herndier, B., et al. (1996) Lytic growth of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) in culture. *Nat Med* 2, 342–346.
34. Sarid, R., Flore, O., Bohenzky, R. A., et al. (1998) Transcription mapping of the Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) genome in a body cavity-based lymphoma cell line (BC-1). *J Virol* 72, 1005–1012.
35. Krishnan, H. H., Naranatt, P. P., Smith, M. S., et al. (2004) Concurrent expression of latent and a limited number of lytic genes with immune modulation and antiapoptotic function by Kaposi's sarcoma-associated herpesvirus early during infection of primary endothelial and fibroblast cells and subsequent decline of lytic gene expression. *J Virol* 78, 3601–3620.
36. Grundhoff, A., Ganem, D. (2004) Inefficient establishment of KSHV latency suggests an additional role for continued lytic replication in Kaposi sarcoma pathogenesis. *J Clin Invest* 113, 124–136.
37. Chen, J., Ueda, K., Sakakibara, S., et al. (2001) Activation of latent Kaposi's sarcoma-associated herpesvirus by demethylation of the promoter of the lytic transactivator. *Proc Natl Acad Sci USA* 98, 4119–4124.
38. Chang, J., Renne, R., Dittmer, D., et al. (2000) Inflammatory cytokines and the reactivation of Kaposi's sarcoma-associated herpesvirus lytic replication. *Virology* 266, 17–25.
39. Davis, D. A., Rinderknecht, A. S., Zoeteweij, J. P., et al. (2001) Hypoxia induces lytic replication of Kaposi sarcoma-associated herpesvirus. *Blood* 97, 3244–3250.
40. Deutsch, E., Cohen, A., Kazimirsky, G., et al. (2004) Role of protein kinase C delta in reactivation of Kaposi's sarcoma-associated herpesvirus. *J Virol* 78, 10187–10192.
41. Sgarbanti, M., Arguello, M., tenOever, B. R., et al. (2004) A requirement for NF-kappaB induction in the production of replication-competent HHV-8 virions. *Oncogene* 23, 5770–5780.
42. McAllister, S. C., Hansen, S. G., Messaoudi, I., et al. (2005) Increased efficiency of phorbol ester-induced lytic reactivation of Kaposi's sarcoma-associated herpesvirus during S phase. *J Virol* 79, 2626–2630.
43. Brown, H. J., McBride, W. H., Zack, J. A., et al. (2005) Prostratin and bortezomib are novel inducers of latent Kaposi's sarcoma-associated herpesvirus. *Antivir Ther* 10, 745–751.
44. Chang, M., Brown, H. J., Collado-Hidalgo, A., et al. (2005) Beta-adrenoreceptors reactivate Kaposi's sarcoma-associated herpesvirus lytic replication via PKA-dependent control of viral RTA. *J Virol* 79, 13538–13547.
45. Morris, T. L., Arnold, R. R., Webster-Cyriaque, J. (2007) Signaling cascades triggered by bacterial metabolic end-products during KSHV reactivation. *J Virol* 81, 6032–6042.
46. Staudt, M. R., Dittmer, D. P. (2007) The Rta/Orf50 transactivator proteins of the gamma-herpesviridae. *Curr Top Microbiol Immunol* 312, 71–100.
47. Whitby, D., Marshall, V. A., Bagni, R. K., et al. (2007) Reactivation of Kaposi's sarcoma-associated herpesvirus by natural products from Kaposi's sarcoma endemic regions. *Int J Cancer* 120, 321–328.
48. Aoki, Y., Tosato, G. (2007) Interactions between HIV-1 Tat and KSHV. *Curr Top Microbiol Immunol* 312, 309–326.
49. Parravicini, C., Olsen, S. J., Capra, M., et al. (1997) Risk of Kaposi's sarcoma-associated herpes virus transmission from donor allografts among Italian posttransplant Kaposi's sarcoma patients. *Blood* 90, 2826–2829.
50. Campbell, T. B., Borok, M., Gwanzura, L., et al. (2000) Relationship of human herpesvirus 8 peripheral blood virus load and Kaposi's sarcoma clinical stage. *AIDS* 14, 2109–2116.
51. Quinlivan, E. B., Zhang, C., Stewart, P. W., et al. (2002) Elevated virus loads of Kaposi's sarcoma-associated human herpesvirus 8 predict Kaposi's sarcoma disease progression, but elevated levels of human immunodeficiency virus type 1 do not. *J Infect Dis* 185, 1736–1744.
52. Tedeschi, R., Enbom, M., Bidoli, E., et al. (2001) Viral load of human herpesvirus 8 in

peripheral blood of human immunodeficiency virus-infected patients with Kaposi's sarcoma. *J Clin Microbiol* 39, 4269–4273.

53. Polstra, A. M., Cornelissen, M., Goudsmit, J., et al. (2004) Retrospective, longitudinal analysis of serum human herpesvirus-8 viral DNA load in AIDS-related Kaposi's sarcoma patients before and after diagnosis. *J Med Virol* 74, 390–396.
54. Pellet, C., Chevret, S., Frances, C., et al. (2002) Prognostic value of quantitative Kaposi sarcoma-associated herpesvirus load in posttransplantation Kaposi sarcoma. *J Infect Dis* 186, 110–113.
55. Gessain, A., Duprez, R. (2005) Spindle cells and their role in Kaposi's sarcoma. *Int J Biochem Cell Biol* 37, 2457–2465.
56. Montaner, S., Sodhi, A., Molinolo, A., et al. (2003) Endothelial infection with KSHV genes in vivo reveals that vGPCR initiates Kaposi's sarcomagenesis and can promote the tumorigenic potential of viral latent genes. *Cancer Cell* 3, 23–36.
57. Guo, H. G., Sadowska, M., Reid, W., et al. (2003) Kaposi's sarcoma-like tumors in a human herpesvirus 8 ORF74 transgenic mouse. *J Virol* 77, 2631–2639.
58. Grisotto, M. G., Garin, A., Martin, A. P., et al. (2006) The human herpesvirus 8 chemokine receptor vGPCR triggers autonomous proliferation of endothelial cells. *J Clin Invest* 116, 1264–1273.
59. Mutlu, A. D., Cavallin, L. E., Vincent, L., et al. (2007) In vivo-restricted and reversible malignancy induced by human herpesvirus-8 KSHV: a cell and animal model of virally induced Kaposi's sarcoma. *Cancer Cell* 11, 245–258.
60. Brown, E. E., Whitby, D., Vitale, F., et al. (2006) Virologic, hematologic, and immunologic risk factors for classic Kaposi sarcoma. *Cancer* 107, 2282–2290.
61. Hengge, U. R., Ruzicka, T., Tyring, S. K., et al. (2002) Update on Kaposi's sarcoma and other HHV8 associated diseases. Part 2: pathogenesis, Castleman's disease, and pleural effusion lymphoma. *Lancet Infect Dis* 2, 344–352.
62. Cesarman, E., Mesri, E. A. (2007) Kaposi sarcoma-associated herpesvirus and other viruses in human lymphomagenesis. *Curr Top Microbiol Immunol* 312, 263–287.
63. Casper, C. (2005) The aetiology and management of Castleman disease at 50 years: translating pathophysiology to patient care. *Br J Haematol* 129, 3–17.
64. Whitby, D., Howard, M. R., Tenant-Flowers, M., et al. (1995) Detection of Kaposi sarcoma associated herpesvirus in peripheral blood of HIV-infected individuals and progression to Kaposi's sarcoma. *Lancet* 346, 799–802.
65. Widmer, I. C., Erb, P., Grob, H., et al. (2006) Human herpesvirus 8 oral shedding in HIV-infected men with and without Kaposi sarcoma. J *Acquir Immune Defic Syndr* 42, 420–425.
66. Boldogh, I., Szaniszlo, P., Bresnahan, W. A., et al. (1996) Kaposi's sarcoma herpesvirus-like DNA sequences in the saliva of individuals infected with human immunodeficiency virus. *Clin Infect Dis* 23, 406–407.
67. Blackbourn, D. J., Lennette, E. T., Ambroziak, J., et al. (1998) Human herpesvirus 8 detection in nasal secretions and saliva. *J Infect Dis* 177, 213–216.
68. Pauk, J., Huang, M. L., Brodie, S. J., et al. (2000) Mucosal shedding of human herpesvirus 8 in men. *N Engl J Med* 343, 1369–1377.
69. Pellett, P. E., Spira, T. J., Bagasra, O., et al. (1999) Multicenter comparison of PCR assays for detection of human herpesvirus 8 DNA in semen. *J Clin Microbiol* 37, 1298–1301.
70. Diamond, C., Brodie, S. J., Krieger, J. N., et al. (1998) Human herpesvirus 8 in the prostate glands of men with Kaposi's sarcoma. *J Virol* 72, 6223–6227.
71. Cerimele, F., Curreli, F., Ely, S., et al. (2001) Kaposi's sarcoma-associated herpesvirus can productively infect primary human keratinocytes and alter their growth properties. *J Virol* 75, 2435–2443.
72. Akula, S. M., Wang, F. Z., Vieira, J., et al. (2001) Human herpesvirus 8 interaction with target cells involves heparan sulfate. *Virology* 282, 245–255.
73. Krishnan, H. H., Sharma-Walia, N., Zeng, L., et al. (2005) Envelope glycoprotein gB of Kaposi's sarcoma-associated herpesvirus is essential for egress from infected cells. *J Virol* 79, 10952–10967.
74. Martin, J. N., Ganem, D. E., Osmond, D. H., et al. (1998) Sexual transmission and the natural history of human herpesvirus 8 infection. *N Engl J Med* 338, 948–954.
75. Blackbourn, D. J., Osmond, D., Levy, J. A., et al. (1999) Increased human herpesvirus 8 seroprevalence in young homosexual men who have multiple sex contacts with different partners. *J Infect Dis* 179, 237–239.
76. Grulich, A. E., Kaldor, J. M., Hendry, O., et al. (1997) Risk of Kaposi's sarcoma and oroanal sexual contact. *Am J Epidemiol* 145, 673–679.
77. Smith, N. A., Sabin, C. A., Gopal, R., et al. (1999) Serologic evidence of human herpesvirus 8 transmission by homosexual but not heterosexual sex. *J Infect Dis* 180, 600–606.
78. Osmond, D. H., Buchbinder, S., Cheng, A., et al. (2002) Prevalence of Kaposi sarcoma-associated herpesvirus infection in homosexual men at beginning of and during the HIV epidemic. *JAMA* 287, 221–225.

79. Mayama, S., Cuevas ,L. E., Sheldon, J., et al. (1998) Prevalence and transmission of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) in Ugandan children and adolescents. *Int J Cancer* 77, 817–820.
80. Gessain, A., Mauclere, P., van Beveren, M., et al. (1999) Human herpesvirus 8 primary infection occurs during childhood in Cameroon, Central Africa. *Int J Cancer* 81, 189–192.
81. Bourboulia, D., Whitby, D., Boshoff, C., et al. (1998) Serologic evidence for mother-to-child transmission of Kaposi sarcoma- associated herpesvirus infection [letter]. *JAMA* 280, 31–32.
82. Plancoulaine, S., Abel, L., van Beveren, M., et al. (2000) Human herpesvirus 8 transmission from mother to child and between siblings in an endemic population. *Lancet* 356, 1062–1065.
83. Sarmati, L., Carlo, T., Rossella, S., et al. (2004) Human herpesvirus-8 infection in pregnancy and labor: lack of evidence of vertical transmission. *J Med Virol* 72, 462–466.
84. Angeloni, A., Heston, L., Uccini, S., et al. (1998) High prevalence of antibodies to human herpesvirus 8 in relatives of patients with classic Kaposi's sarcoma from Sardinia. *J Infect Dis* 177, 1715–1718.
85. Guttman-Yassky, E., Kra-Oz, Z., Dubnov, J., et al. (2005) Infection with Kaposi's sarcoma-associated herpesvirus among families of patients with classic Kaposi's sarcoma. *Arch Dermatol* 141, 1429–1434.
86. Davidovici, B., Karakis, I., Bourboulia, D., et al. (2001) Seroepidemiology and molecular epidemiology of Kaposi's Sarcoma-associated herpesvirus among Jewish population groups in Israel. *J Natl Cancer Inst* 93, 194–202.
87. Plancoulaine, S., Abel, L., Tregouet, D., et al. (2004) Respective roles of serological status and blood specific antihuman herpesvirus 8 antibody levels in human herpesvirus 8 intrafamilial transmission in a highly endemic area. *Cancer Res* 64, 8782–8787.
88. Mbulaiteye, S. M., Biggar, R. J., Bakaki, P. M., et al. (2003) Human herpesvirus 8 infection and transfusion history in children with sickle-cell disease in Uganda. *J Natl Cancer Inst* 95, 1330–1335.
89. Hladik, W., Dollard, S. C., Mermin, J., et al. (2006) Transmission of human herpesvirus 8 by blood transfusion. *N Engl J Med* 355, 1331–1338.
90. Barozzi, P., Luppi, M., Facchetti, F., et al. (2003) Post-transplant Kaposi sarcoma originates from the seeding of donor-derived progenitors. *Nat Med* 9, 554–561.
91. Parkin, D. M., Whelan, S. L., Ferlay, J., et al. (2003) *Cancer Incidence in Five Continents*, Vol. VIII. IARC Scientific Publications, Lyon, France.
92. Gao, S. J., Kingsley, L., Li, M., et al. (1996) KSHV antibodies among Americans, Italians and Ugandans with and without Kaposi's sarcoma. *Nat Med* 2, 925–928.
93. Gao, S. J., Kingsley, L., Hoover, D. R., et al. (1996) Seroconversion to antibodies against Kaposi's sarcoma-associated herpesvirus-related latent nuclear antigens before the development of Kaposi's sarcoma. *N Engl J Med* 335, 233–241.
94. Kedes, D. H., Operskalski, E., Busch, M., et al. (1996) The seroepidemiology of human herpesvirus 8 (Kaposi's sarcoma-associated herpesvirus): distribution of infection in KS risk groups and evidence for sexual transmission. *Nat Med* 2, 918–924.
95. Simpson, G. R., Schulz, T. F., Whitby, D., et al. (1996) Prevalence of Kaposi's sarcoma associated herpesvirus infection measured by antibodies to recombinant capsid protein and latent immunofluorescence antigen. *Lancet* 348, 1133–1138.
96. Laney, A. S., Peters, J. S., Manzi, S. M., et al. (2006) Use of a multiantigen detection algorithm for diagnosis of Kaposi's sarcoma-associated herpesvirus infection. *J Clin Microbiol* 44, 3734–3741.
97. Dukers, N. H., Rezza, G. (2003) Human herpesvirus 8 epidemiology: what we do and do not know. *AIDS* 17, 1717–1730.
98. Mohanna, S., Maco, V., Bravo, F., et al. (2005) Epidemiology and clinical characteristics of classic Kaposi's sarcoma, seroprevalence, and variants of human herpesvirus 8 in South America: a critical review of an old disease. *Int J Infect Dis* 9, 239–250.
99. Dedicoat, M., Newton, R. (2003) Review of the distribution of Kaposi's sarcoma-associated herpesvirus (KSHV) in Africa in relation to the incidence of Kaposi's sarcoma. *Br J Cancer* 88, 1–3.
100. Plancoulaine, S., Gessain, A., van Beveren, M., et al. (2003) Evidence for a recessive major gene predisposing to human herpesvirus 8 (HHV-8) infection in a population in which HHV-8 is endemic. *J Infect Dis* 187, 1944–1950.
101. Ariyoshi, K., Schim, V. D. L., Cook, P., et al. (1998) Kaposi's sarcoma in the Gambia, West Africa is less frequent in human immunodeficiency virus type 2 than in human immunodeficiency virus type 1 infection despite a high prevalence of human herpesvirus 8. *J Hum Virol* 1, 193–199.
102. He, F., Wang, X., He, B., et al. (2007) Human herpesvirus 8: serovprevalence and correlates in tumor patients from Xinjiang, China. *J Med Virol* 79, 161–166.

103. Biggar, R. J., Whitby, D., Marshall, V., et al. (2000) Human herpesvirus 8 in Brazilian Amerindians: a hyperendemic population with a new subtype. *J Infect Dis* 181, 1562–1568.
104. Grossman, Z., Iscovich, J., Schwartz, F., et al. (2002) Absence of Kaposi sarcoma among Ethiopian immigrants to Israel despite high seroprevalence of human herpesvirus 8. *Mayo Clin Proc* 77, 905–909.
105. Cottoni, E., Masia, I. M., Masala, M. V., et al. (1996) Familial Kaposi's sarcoma: case reports and review of the literature. *Acta Derm Venereol* 76, 59–61.
106. Guttman-Yassky, E., Cohen, A., Kra-Oz, Z., et al. (2004) Familial clustering of classic Kaposi sarcoma. *J Infect Dis* 189, 2023–2026.
107. Hayward, G. S., Zong, J. C. (2007) Modern evolutionary history of the human KSHV genome. *Curr Top Microbiol Immunol* 312, 1–42.
108. Sarid, R., Olsen, S. J., Moore, P. S. (1999) Kaposi's sarcoma-associated herpesvirus epidemiology, virology, and molecular biology. *Adv Virus Res* 52, 139–232.
109. Vitale, F., Briffa, D. V., Whitby, D., et al. (2001) Kaposi's sarcoma herpes virus and Kaposi's sarcoma in the elderly populations of 3 Mediterranean islands. *Int J Cancer* 91, 588–591.
110. Woodle, E. S., Hanaway, M., Buell, J., et al. (2001) Kaposi sarcoma: an analysis of the US and international experiences from the Israel Penn International Transplant Tumor Registry. *Transplant Proc* 33, 3660–3661.
111. Renwick, N., Halaby, T., Weverling, G. J., et al. (1998) Seroconversion for human herpesvirus 8 during HIV infection is highly predictive of Kaposi's sarcoma. *AIDS* 12, 2481–2488.
112. Dorak, M. T., Yee, L. J., Tang, J., et al. (2005) HLA-B, -DRB1/3/4/5, and -DQB1 gene polymorphisms in human immunodeficiency virus-related Kaposi's sarcoma. *J Med Virol* 76, 302–310.
113. Gaya, A., Esteve, A., Casabona, J., et al. (2004) Amino acid residue at position 13 in HLA-DR beta chain plays a critical role in the development of Kaposi's sarcoma in AIDS patients. *AIDS* 18, 199–204.
114. Cottoni, F., Masala, M. V., Santarelli, R., et al. (2004) Susceptibility to human herpesvirus-8 infection in a healthy population from Sardinia is not directly correlated with the expression of HLA-DR alleles. *Br J Dermatol* 151, 247–249.
115. Foster, C. B., Lehrnbecher, T., Samuels, S., et al. (2000) An IL6 promoter polymorphism is associated with a lifetime risk of development of Kaposi sarcoma in men infected with human immunodeficiency virus. *Blood* 96, 2562–2567.
116. Lehrnbecher, T. L., Foster, C. B., Zhu, S., et al. (2000) Variant genotypes of FcgammaRIIIA influence the development of Kaposi's sarcoma in HIV-infected men. *Blood* 95, 2386–2390.
117. Brown, E. E., Fallin, D., Ruczinski, I., et al. (2006) Associations of classic Kaposi sarcoma with common variants in genes that modulate host immunity. *Cancer Epidemiol Biomarkers Prev* 15, 926–934.
118. Simonart, T. (2006) Role of environmental factors in the pathogenesis of classic and African-endemic Kaposi sarcoma. *Cancer Lett* 244, 1–7.
119. Montella, M., Serraino, D., Crispo, A., et al. (2000) Is volcanic soil a cofactor for classic Kaposi's sarcoma? *Eur J Epidemiol* 16, 1185–1186.
120. Nawar, E., Mbulaiteye, S. M., Gallant, J. E., et al. (2005) Risk factors for Kaposi's sarcoma among HHV-8 seropositive homosexual men with AIDS. *Int J Cancer* 115, 296–300.
121. Goedert, J. J., Vitale, F., Lauria, C., et al. (2002) Risk factors for classical Kaposi's sarcoma. *J Natl Cancer Inst* 94, 1712–1718.
122. Guttman-Yassky, E., Dubnov, J., Kra-Oz, Z., et al. (2006) Classic Kaposi sarcoma. Which KSHV-seropositive individuals are at risk? *Cancer* 106, 413–419.
123. Pfeffer, S., Sewer, A., Lagos-Quintana, M., et al. (2005) Identification of microRNAs of the herpesvirus family. *Nat Methods* 2, 269–276.
124. Cai, X., Lu, S., Zhang, Z., et al. (2005) Kaposi's sarcoma-associated herpesvirus expresses an array of viral microRNAs in latently infected cells. *Proc Natl Acad Sci USA* 102, 5570–5575.
125. Samols, M. A., Hu, J., Skalsky, R. L., et al. (2005) Cloning and identification of a microRNA cluster within the latency-associated region of Kaposi's sarcoma-associated herpesvirus. *J Virol* 79, 9301–9305.
126. Grundhoff, A., Sullivan, C. S., Ganem, D. (2006) A combined computational and microarray-based approach identifies novel microRNAs encoded by human gamma-herpesviruses. *RNA* 12, 733–750.
127. Moore, P. S., Chang, Y. (2003) Kaposi's sarcoma-associated herpesvirus immunoevasion and tumorigenesis: two sides of the same coin? *Annu Rev Microbiol* 57, 609–639.
128. Schulz, T. F. (2006) The pleiotropic effects of Kaposi's sarcoma herpesvirus. *J Pathol* 208, 187–198.
129. Jarviluoma, A., Ojala, P. M. (2006) Cell signaling pathways engaged by KSHV. *Biochim Biophys Acta* 1766, 140–158.
130. Rezaee, S. A., Cunningham, C., Davison, A. J., et al. (2006) Kaposi's sarcoma-associated herpesvirus immune modulation: an overview. *J Gen Virol* 87, 1781–1804.
131. Sodhi, A., Montaner, S., Gutkind, J. S. (2004) Does dysregulated expression of a deregulated viral GPCR trigger Kaposi's sarcomagenesis? *FASEB J* 18, 422–427.
132. Boshoff, C. (2003) Kaposi virus scores cancer coup. *Nat Med* 9, 261–262.

Chapter 7

Thriving for Clues in Variations Seen in Mortality and Incidence of Cancer: Geographic Patterns, Temporal Trends, and Human Population Diversities in Cancer Incidence and Mortality

Alireza Mosavi-Jarrahi and Mohammad Ali Mohagheghi

Summary

To address the diverse features of the cancer epidemic, scientists have continually been seeking out different aspects of the cancer problem from cell proliferation at the microenvironment of the cancerous cells to large-scale diversities and variations seen in the incidence of cancer and prevalence of exposure to potential or known carcinogens in as many global locations and as many occasions as possible. The variations of cancer occurrence in the three dimensions of place, time, and human population have been the pillar of hypothesis formation in understanding the diverse features of the cancer epidemic. The aim of this chapter is to address the methodologic issues pertaining to changing of cancer incidence and mortality in the three dimensions. For each dimension; a brief description of available information on the cancer incidence and mortality is presented, the sources of available data are introduced, the indicators to be used and their interpretabilities are addressed, and the use of new technologies pertaining to that dimension are discussed.

Key words: Temporal trend, geographical pattern, disease mapping, geographic information system, spatial cluster, migrant studies.

1. Introduction

To address the diverse features of the cancer epidemic, scientists have continually been seeking out different aspects of the cancer problem from cell proliferation at the microenvironment of the cancerous cells to large-scale diversities and variations seen in the incidence of cancer and prevalence of exposure to potential or known carcinogens in as many global locations and as many occasions as possible. Although

Mukish Verma (ed.), *Methods in Molecular Biology, Cancer Epidemiology, vol. 471*
© 2009 Humana Press, a part of Springer Science + Business Media, Totowa, NJ
Book doi: 10.1007/978-1-59745-416-2

striking advances in cellular and molecular technology dealing with cancer at the cellular level have brought new frontiers in understanding the natural history of cancer from its cellular conception to the end of life issues, the notion of discoveries of new frontiers in fighting cancer at both the micro- and macroenvironment relays the development of new thoughts, ideas, and in scientific terms, hypotheses. The starting point of scientific discovery is the formation of a hypothesis generated by the observation of a series of events and the examination of these events against the established body of scientific facts to advance the frontier of understanding a biological or physiological phenomenon.

Epidemiology, as the science of population and health at the macroenvironment level, develops and challenges hypotheses for all aspects of public health issues, including cancer. Descriptive epidemiology, the portrayal of disease or any heath event in the three dimensions of place, time, and human population, has been the pillar of hypothesis formation ever since John Snow hypothesized that cholera is a waterborne disease *(1)*.

Cancer is a biological disease with an especially long latent period and chronic nature. With its complex natural history, from the malignant formation at the cellular level to full-blown symptoms at the organism level, and further complexity at the public health and community levels, the etiology of cancer is a subject of interest to the epidemiologist. Examining incidence variations within the triad of epidemiology (place, time, and person), the epidemiologist must consider the geography and chronology as well as the race, ethnicity, and culture of the human population as they relate to disease. The main purpose of scrutinizing all these variations is to understand the very basic issue of how genes, environment, or both affect the changing pattern of cancer incidence at the community level. The tools of demonstrating the variation in the pattern of cancer occurrence developed into a surveillance system or "cancer registry" as early as the 1920s and 1930s when the first cancer reporting system was established in Germany *(2)*. The cancer registry, now a well-developed surveillance method, systematically collects cancer information either at hospitals (hospital-based cancer registries) or in defined populations (population-based cancer registries) *(3)*. The hospital-based cancer registry, although serving as a mechanism to evaluate cancer patient treatment regimens, generates valuable information for population-based cancer registries. The population-based cancer registry, by recording all cancer cases for a defined population over well-defined geographic boundaries, is serving the public health authorities by providing timely information about the changes in cancer incidence at the population level *(4)*. Although the cancer registry has been a traditional means of collecting cancer information, with the advances in information

technology along with advances in new techniques in cellular molecular biology, new data repositories have been developed and made available to scientists and epidemiologists interested in understanding the etiology and complexity of cancer *(5–7)*. This chapter explores the tools that epidemiologists use to develop hypotheses that aid in the understanding of the etiology and natural history of cancer, thereby addressing the needs of cancer epidemic in the world. The topics discussed in this chapter cover how the triad of epidemiology is used with regard to variations in cancer occurrence, the complexities pertaining to these variations, and the methodological tools available to address such variations.

2. Time: Temporal Variations in Cancer

The practice of examining time trends in cancer mortality and incidence has already been widely and well used *(7–10)*. The total cancer incidence and mortality in the United States has risen slowly until recently when both indicators saw a slight decline *(11)*. Between 1975 and 1982, the age-adjusted rate of total cancer mortality rose by 1%, for both male and female. The major contributor to this rise was the increase in lung cancer deaths among both male and female *(12,13)*. Breast cancer, ranging in incidence from 27/100,000 in Asian countries to 97/100,000 among Caucasian women in the United States has been one of the most studied cancers in terms of chronological changes in incidence and mortality *(14)*. A steady rise in the global incidence and mortality of breast cancer over past 30 years has been documented, especially with regard to countries with the highest and the lowest rates *(15)*. Breast cancer incidence rose 30–40% from the 1970s to the 1990s in most countries, with the most striking increase among women aged <50 years *(14)*. Very recently, an increase in childhood cancer has been documented in Europe *(16,17)*. The rate of esophageal cancer has been declining in countries with high risk and high rates *(18)*.

There are ample studies regarding changes in the incidence of cancer over time. That, by the passage of time, the incidence and mortality are changing indicates the existence of underlying causes. The identification of such underlying causes is intensely studied by epidemiologists and public health scientists; in the case of an increase in incidence and mortality, so that they may identify preventive measures; or learn the underlying cause and how to apply it if the change is a decrease. The way to address changes over time and draw conclusions on possible new etiologic factors demands its own methodology, indicators and interpretation.

2.1. Indicators of Temporal Changes

There are four indicators, each with its own pros and cons, available to the epidemiologist to monitor temporal changes in cancer occurrence over time: 1) incidence, 2) mortality, 3) prevalence, and 4) proportionate mortality and morbidity.

1. The incidence is a true measure of disease frequency that has very high value for the etiologic investigation of any disease, especially cancer; however, the availability of information regarding the incidence over time may be lacking, as is the case in most developing countries *(3)*. The cancer epidemiologist obtains incidences and rates of various cancers based on data generated by cancer registries. Given that a high-quality cancer registry is a serious undertaking and that registries usually operate in small populations or samples of populations, the availability of incidence data for an entire population is not always available to monitor temporal changes *(19)*. Regardless, incidence is in fact the best indicator to monitor changes of cancer frequency over time. When studied over time, the use of incidence data may have certain limitations, which are discussed later in subheading 2.3.
2. Another widely used indicator in the evaluation of variations over time is mortality rates. Because mortality is registered in almost every society, the mortality rate has been a very widely used indicator to monitor almost all disease chronologically *(20)*. Mortality rates are, however, subject to certain limitations *(20–23)*, some of which are temporal and often related to calendar times *(24,25)*. Because mortality rates are based on the cause of death as stated in the death certificate and they are subject to systematic errors *(20)*, mortality rate is not as good an indicator as incidence for monitoring temporal changes. Mortality rates are highly dependent on the classification used by the death registry office *(26,27)*, and on the practice of cause-of-death interpretation and establishment by the clinician issuing the death certificate *(28,29)*. The errors and inconsistencies inherent in mortality data must be addressed when mortality is used as an indicator for examining temporal changes in disease.
3. The prevalence rates from surveys repeated over the years have been widely used to measure the temporal changes in disease and health events *(30)*. Although there are no actual surveys to measure frequency of cancer, the prevalence of cancer has often been measured or estimated over time by using mortality and survival data *(31–33)*, or it has been estimated from data generated through screening programs for site-specific cancers, when a population screening program can be achieved, such as with breast *(34)* or cervical cancer *(35,36)*. In addition, prevalence measure has been a very good

indicator of risk factors for chronic diseases, especially cancer. Currently, repeated national surveys of lifestyle, nutrition, and smoking in addition to preconditions for certain diseases are continually underway in most of the countries. These repeated surveys and prevalence measures are subject to the same comparability and validity issues that pertain to incidence, and to some extent, mortality data. With the advances in technology, as the sensitivity of tools of measurement are changing and improving, the classification and definition of measurements also are changing, and these changes need to be considered to prevent misinterpretation when changes of prevalence of an event are studied. A very important issue in examining changes in prevalence is the confounding effects of birth-cohort or secular trends over a long period, generally involving years or decades. Such confounding effects must be considered when interpreting changes in prevalence for any indicators of any health event *(37,38)*. Even with the limitations inherent in the study of prevalence measures, however, the changes in prevalence of cancer has been a valuable tool used by epidemiologists to generate hypotheses in the hunt for etiologies to develop new means for the delivery of care to patients and improving the effectiveness of certain interventions in the prevention of cancer *(39)*.

4. Another indicator for the measure of chronologic changes is the proportionate morbidity or proportionate mortality (PM), the proportion of illnesses or deaths over a period that are attributed to various causes in a defined population of all illnesses or deaths. In the absence of a precise denominator for determining morbidity or mortality rates (i.e., the total population in which the illnesses or deaths occurred), PM data have been cautiously used to monitor the changes of disease frequencies over time. PM has severe limitations in interpretation and it is subject to a bias that arises due to changes in relative frequency, as a result of changes in another disease included as part of the denominator. In the absence of incidence and mortality measures, the mortality odds ratio (MOR), suggested by Miettinen and Wang *(40)*, has been proposed as an alternative method to PM *(4)*. The MOR uses the morbidity of the cancer of interest as "case," with another cancer or disease morbidity as "control,", and for the purpose of studying variation over time, the calendar year as the time of exposure. A regression line is fitted to the morbidity odds ratios over the years, and the slope of the regression line is the indicator of changes over time. This method has been used to examine the temporal trends of gastrointestinal cancer in the Iranian population *(41)*.

2.2. Measuring Magnitude of Temporal Changes over Time

Two ways to measure temporal changes of incidence and mortality over a period are graphical and numerical. In the graphical presentation, the incidence or mortality rates is normally graphed against yearly time intervals. Although such a graph is a crude presentation of the temporal change, it conveys a good degree of visualization. To measure the magnitude of change over a period, a direct adjustment called the standard incidence ratio (SIR), or an indirect adjustment called the standard mortality ratio (SMR), can be constructed over units of time in two ways. The SIR or SMR compares incidence or mortality of each year with the previous year; the SIRs and the confidence interval are then graphed against the time interval. Calculation of the SIR or the SMR over time gives smoothed data and further analysis can be performed by fitting a linear regression line to give the slope as a measure of increase or decrease. However, care should be taken in interpreting the slope of the regression line as measures of temporal change because the temporality of calendar time may be lost; calendar time is more informative than time as a continuous variable. Analogous to SIR and SMR, all other indicators, such as prevalence ratio (proportionate mortality ratio; PMR) and MOR, can be used in the same way to demonstrate the magnitude and better visualize temporal changes.

2.3. Validity Issues in Measuring Temporal Changes

Several validity and comparability issues must be considered in the study of trends and temporal changes. 1) With the advancement of technology, diagnostic sensitivity increases, which can introduce a bias in the interpretation of changes *(42)*. This issue is exemplified by the comparison of the incidence by using diagnosis of esophageal cancer based on radiology as it was practiced 30 years ago in the high-risk area of the Caspian littoral, during which >70% of the cases were identified by symptomatic dysphasia or simple radiography of the esophagus, to incidence using diagnosis based on endoscopic biopsy, a modern practice *(18)*. 2) The practice of data collection normally changes with the arrival of new technology. For example, many cancer registries currently use automated reporting versus the manual data collection of previous years that was more subject to the omission of certain cases. Recent studies of cancer registries' validity indicate a differential rate of bases of diagnosis over time, with a tendency to have higher rate of death-certificate-only diagnosis in older times compared in recent times *(43,44)*. 3) The notion of birth and cohort effects should be considered when evaluating temporal trends; birth cohorts may experience differential rates of cancer incidence with the aging of that birth cohort over calendar time *(45)*.

2.4. Interpreting Temporal Changes in Rates

Any change in the rate of cancer, either crude or site specific, must be examined carefully. A change in crude rate may be misleading and needs to be adjusted for confounder and masking variables. Changes have to be grouped based on topologic sites,

because a change in the rate of a subsite cancer may have different etiologic attributes, which is the case for esophageal cancer because cancer in upper part of the esophagus may infer certain risks, whereas another risk is associated with cancer in the lower part of the esophagus. The same is true for various morphologies; cancers arising from the same organ with different morphologies have different risk profiles. For example, adenocarcinoma of the esophagus is basically associated with Barrett's esophagus and gastroesophageal reflux disease syndrome, whereas squamous cell carcinoma of the esophagus is more often associated with as risk factors, such as the use of alcohol and smoking.

In examining changes, all possible artifacts must be excluded before a genuine hypothesis is concluded or an etiologic risk is attributed to the change. The changes in temporal rates have been of great value in developing hypotheses in understanding etiologic factors. Any temporal changes as are considered to be the beginning of a collective effort, and such an effort requires information on the established risk factors and their frequencies or their changes over time to enable an epidemiologist to draw hypotheses for further examination or even make conclusions.

3. Place: Geographical Variations in Cancer

Place, the second of the triad of descriptive epidemiology, is used to indicate a mixture of lifestyle, physical environment, and genetic factors (common genetic pool). Although there is no exact definition of what a place is, the political boundaries at the international level and the administrative boundaries at the national level make the core practicalities of a place where disease variation can be measured or described. The disease incidence is mainly measured on the administrative boundaries where population counts are available. The purpose of any study exploring the variation of cancer incidence according to place (or, as it is loosely called, geographic variation) is to identify the possible causes to explain the variation. In cancer epidemiology, the study of geographic variation has a broad use, and it is done in different scales, from variation seen at the continent level, to variation seen in urban areas and clusters of cancer cases around some sources of hazards *(46–49)*.

The introduction of information technology and use of geographic information systems (GISs) have provided valuable opportunities in describing these geographic variations and furthering the development of hypotheses regarding the etiology of cancer *(50)*. Geographic variation in cancer occurrence is seen in almost all geographic scales. There are distinct patterns of

site-specific cancer occurrence in the developed versus the developing countries. Cancer in certain sites, such as the stomach and esophagus *(51)*, are generally more common in the low-resource countries in Asia (except for Japan, which has a high incidence of stomach cancer, although it is considered a developed country); cervix in Latin America and the Indian subcontinent; and cancers of breast, prostate, and colon are more common in the affluent countries of North America and Europe *(51,52)*. Along with the variation seen on scales as large as continents and countries, there is variation in cancer occurrence within a country and sometimes in distances as small as 100 km. Such small places with distinct patterns of cancer, termed as hot spots, are seen in Europe, exemplified in esophageal cancer in France or Iran, or other parts of world *(53)*. There are very diverse methodologies to study the variation of cancer in different places. Most of these studies are based on the incidence and mortality rates, where incidence is defined using some arbitrary boundaries or census tracts, depending on the availability of information about the population or denominator. New techniques, such as GIS and remote sensing technologies, have been used with complex analytical procedures to draw hypotheses on the etiology of diseases, including cancer across geographic scales *(54)*. The complicated, but existing, technique of cluster detection and investigation is another tool that has been available to epidemiologists to address geographic variations at smaller scales.

3.1. Sources of Data Available to Address Geographic Variation

In addition to several ad hoc resources available to the cancer epidemiologist to address geographic variations, there are two important sources with a vast repository of data. The most useful resource is the Cancer Incidence in Five Continents (CI5), published by the International Agency for Research on Cancer. Currently, with eight volumes published thus far, the CI5 reports the data of cancer registries from around the world that have met the high standards of completeness and validity. Because the CI5 publication is centrally managed, its published data are comparable, making it a very valuable resource for epidemiologists interested in geographical variation in cancer incidence across the globe. The CI5 provides incidence data for >30 site-specific cancers by age group and sex *(55)*.

The second main source of this type of cancer information is the World Health Organization (WHO) mortality database. This database also has the benefit of comparability in terms of the coding system used, because mortality around the world is reported using different ICD coding versions. However, mortality data is always subject to issues of validity at the point of establishment of cause of death in the death certificate. The practice of establishing cause of death varies from country to country, depending on autopsy rate and other influencing factors. In addition, because

mortality data are highly dependent on the quality of health care delivery and screening programs, interpretation of the mortality data across the different geographical entities could be misleading, especially when drawing conclusions on the etiologic aspect of cancer.

3.2. Tools Available to Explore Geographical Variation of Cancer

3.2.1. Disease Mapping

Maps have long been a tool to describe the distribution of disease based on geographical scales. Although the very first epidemiologic investigation of cholera by Jon Snow is based on the geographic distribution of mortality from cholera, the first maps or atlases of disease frequency were published in 1930, describing the geographic variation in cancer mortality in England and Wales *(56)*. A survey of resources in disease mapping in 1991 revealed approximately 49 international, national, and regional disease atlases *(57)*.

Maps convey instant visual information on the spatial distribution of diseases and can identify subtle patterns, which may be missed in tabular presentation. Their purpose is usually to display variation in the occurrence of cancer morbidity or mortality mainly related to underlying sociodemographic indicators. However, the map is the basic tool of visualization used to formulate hypotheses ranging from the disease etiology to the effect of service delivery, survival as well as mortality *(58)*.

Although disease mapping has the benefit of visualizing differences in the distribution of cancer across a given area of interest, the degree of differentiation is based on visual perception, which is influenced by various features of the map such as the plotting symbols used (types of shading, simple boundary maps, choroplething, and so forth). Considerable caution is required regarding avid overinterpretation when dealing with disease maps. An empirical study has found that the way a map presents data has much more effect on observer perception of the spatial variation as would the actual differences existing in the data *(59)*. Another important issue in mapping cancer data is the choice of summary measure presented and visualized in the maps. Regardless, because maps and visualization carry the main notion of comparing some measure of disease frequency, caution must be taken regarding the choice of summary measure, two categories of which are available for mapping cancer variation. The first category is descriptive measures of frequency, incidence, mortality, prevalence, and relative frequency, and the second category includes measures of associations such as ISR, SMR, PMR, and MOR. When a measure of association is used to map cancer data, each segment of the map is presented as a ratio of a standardized rate in that segment to a reference rate in a population or place, and the choice of reference population has to be carefully considered, especially when the variation across the mapping area is not very large.

3.2.2. Geographic Information Systems

GISs capture, store, retrieve, analyze, and display spatial data in an automated manner *(60)*. GISs treat spatial data as unique, because they are linked to a geographic map. The main components of a GIS, in addition to a database, include spatial or map information and tools that link and perform spatial analyses of disease or any health-related event. The GIS provide unique tools to perform three general types of spatial analysis tasks: visualization, exploratory data analysis, and model building *(61)*.

Used in new ways to explore data from traditional statistical analyses, visualization shows variation in values over an area by the locations of outlier and influential values on maps, thereby enhancing epidemiologic research. Although such tools are being developed and explored, they would benefit greatly from a closer and more seamless link between the statistical packages and GIS. Visualization has been a basic tool of presenting and summarizing data either through a simple graph or complex and overlying maps, the pro and cons of which were discussed in subheading 3.2.1.

As the second class of GIS methods, exploratory spatial analysis lets the analyst intelligently search the spatial data, identify spatial patterns of interest, especially spatial clusters of disease, and prepare hypotheses to direct research.

The third class of spatial analysis, modeling, uses procedures that test hypotheses regarding the etiology and transmission of disease. By integrating standard statistical and epidemiologic techniques, the GIS produces data to use in epidemiologic models. Statistical analysis results are shown, and modeling processes can then be displayed over space.

GIS technology and geospatial tools have been used at the United States' National Cancer Institute to 1) identify and display the geographic patterns of cancer incidence and mortality rates in the United States and their change over time; 2) create complex databases to study cancer screening, diagnosis, and survival at the community level; 3) assess environmental exposure using satellite imagery; 4) model spatial statistics to estimate cancer incidence, prevalence, and survival for each U.S. state; 5) communicate local cancer information to public health professionals and the public at large using interactive Web-based tools; 6) identify health disparities at the local level by comparing cancer results across demographic subgroups; and 7) develop new methods of presenting geospatial data to communicate unambiguously to the public and to allow researchers to examine complex multivariate data.

GIS technology has opened a new window to the art of exploratory data analysis and visualizing the differences in cancer frequency rates across large geographic areas. However, very often the building blocks of maps and the units of analysis in the GIS are the polygons composed of certain geographic or

administrative boundaries. The measures of cancer frequency rely on the incidence of cancer in a defined population, which in turn depends on the boundaries drawn for administrative purpose, giving rise to enormous limitations in spatial modeling analysis. The limitations of GIS in cancer and other diseases are thoroughly discussed by Jacquez *(62)*.

3.2.3. Cluster Investigations in Space

The term "disease cluster" refers to an excess of cases above a background rate delimited in space and time. In theory, any cluster of disease that cannot be explained with some established and nonmodifiable risk factors (such as age or gender) must be thoroughly investigated to find the underlying cause. When studying the etiology of cancers, the search for cancer clusters is another approach used extensively in the epidemiology of childhood cancer. Studies in highly urbanized areas, such as Liverpool, Greater London, Buffalo City, San Francisco, and metropolitan Atlanta, indicate clustering of childhood cancer in both space and time *(63)*. Furthermore, recent studies of space–time clustering have suggested clustering for lymphoblastic leukemia (precursor B-cell subtype), Wilms' tumor, and soft tissue sarcoma *(64)*.

Although methodological difficulties exist in the investigation of clusters of cancer cases, attempts have been made to address those difficulties, and several well-approached methodologies have been developed. Although most of the clustering investigations have shown no etiologic relationship between the clustered cancer and etiologic factors, the cluster investigation still stands as one of the attempts to understand the etiology behind ascending rates of cancer occurrence in space and time.

To investigate or to detect a cluster, the following are necessary:

- Address the validity of all cases to prevent the inclusion of data from duplicate records and erroneous reports, and ensure the accuracy of the location of each case (problematic in some regions) and a proper incidence date, to make sure recurrent cases are excluded.
- Clearly define the time and space boundaries to construct a population denominator by age and gender, usually from census records.
- Be able to estimate the anticipated numbers of cases based on age- and gender-specific background rates, such as those obtained from published regional or national data.
- Choose the right statistical methodology to investigate the cluster.

How a cluster of cancer cases comes to the attention of the epidemiologist is an issue of great interest, with the availability of electronic cancer databases, plus the fact that the majority of

cases can be identified in space and time, the opportunity to discover clusters of cases is increasing. Although several approaches and methodologies for cluster detection are available, Kulldorff's scan statistic *(65)* has been used as a compelling tool, especially that this technique addresses the multiple comparison problems. This technique has been widely used in the search for clustering and the author has used this technique to investigate childhood cancer in Tehran, resulting in the detection of several clusters, and guiding researchers toward the locations in which to investigate each cluster *(66)*.

3.2.4. Problems When Detecting Cancer Clusters in Space

When investigating any cluster, investigators are faced with three kinds of problems. The first is the choice of analytical methods. Regarding spatial clusters with truly environmental causes, considering that environmental hazard dispersion follows a certain order, if there is a causal relationship between the hazard and the cancer occurrence, the clustering in space should follow the order of the dispersion. Thus, detecting such clusters is complicated by the choice of analytical tool.

Another problem with cluster detection is the choice of the denominator or the population at risk; for any increase in cases to be considered a cluster, there also must be an increase in geographical scale. In cases of geographical scale, this can be achieved where determination of the size of the population at risk is possible. However, as in many cases, the number of people at risk may not be available on a scale to differentiate between an area of increasing rate and the area of normal rate. In many cases, the lack of denominators may, in fact, prevent the detection of clusters, especially in developing countries where resources are scarce and population statistics may not be available.

The third problem is the resolution, or the size of geographic boundaries, at which cluster detection can be performed. Recent advances in data availability have created new opportunities for investigators to study variations in cancer occurrence on a local (small-area) scale. Such investigations may include exposures to local sources of environmental pollution and the distribution of locally varying socioeconomic and behavioral factors. They also present new challenges, because as the scale of the investigation becomes narrowed to a particular small area or group of areas, the reduced size of the population-at-risk leads to small numbers of events, resulting in unstable risk estimates that must be addressed using complex methodologies.

3.3. Interpreting Geographical Variations

Geographical variation seen in cancer occurrence is dependent upon the geographical scale in which the occurrence is observed. Interpretation of the observed variation in the geographical distribution of cancer frequency depends on three main components,

including the geographical scale, the summary measures used in the analysis, and the technology and statistical complexity.

The geographical scale is an issue that brings precision to the interpretation of the observed variation. The differences in the incidence and pattern of cancer occurrence at the continental scale has already established the frequency of cancers at different socioeconomic scales; the cancers of lung, colon, breast, endometrium, and ovary are frequent cancers among relatively rich European countries and North America and the cancers of esophagus, cervix, and stomach are the cancers seen mostly in the relatively poor countries *(52)*. The continental scale is so large that it includes a socioeconomic gradient with certain cancer frequencies across the globe, the interpretation of which is too broad and ambiguous and has very low value from the aspects of cancer control and etiology. Large-scale geographical variation must be interpreted cautiously as such variation could be attributed to a range of factors, such as the differences seen in the patterns of cancer and magnitudes of occurrence between European and African countries that may be attributed to race, ethnicity, socioeconomic level, and lifestyle as well as environmental hazards. Further classification of rates at this large scale may render it impossible to draw conclusions or generate hypotheses and is subject to difficulty and ambiguity. Although it will be very hard to define a very large or very small geographical scale, the smaller the geographical scale the more refined the assessment of the etiologic attributes. For example, variations seen in the incidence of esophageal cancer in Linxian, China, and Gonbad, Iran *(18)*, has resulted in the generation of several hypotheses that have been tested and evaluated to understand the etiology of esophageal cancer. Small-scale geographical entities, which include more homogeneous populations in terms of ethnicity, culture, lifestyle and nutritional habits, make evaluation and assessment of the etiology behind a variation less subject to ambiguity and ecological fallacy.

Another important factor in interpreting geographical variation is the summary measures that are used to describe or scale the magnitude of variations. As mentioned, the incidence measures, when adjusted for influencing factors such as age and gender distribution, are the best measure for assessing geographical variation. Other measures, such as prevalence (PMR and MOR) are subject to certain comparability problems that may distort interpretation and give rise to wrong conclusions.

The last factor in the interpretation geographical variation studies is the use of the new technologies, statistical modeling, and complexities to draw conclusions. One of the techniques used in the study of cancer, especially childhood cancer, has been the investigation of possible case clustering around certain environmental hazards such as landfills or nuclear reactors. The

investigation of a cluster is in fact another way of describing small-scale geographical variation. This depends on a statistical model that incorporates a great deal of assumption. Although the detection of cancer clusters are very important to epidemiologists looking at the etiology of cancer, detected clusters must be examined carefully, and the result of certain clusters requires other means of investigation to address the etiology behind the cluster.

4. Person: Variations in Race and Ethnic Diversities

The third of the epidemiology triad is the human host. The diversity of disease rates among humans as a population has been of prime interest to epidemiologists in the study of etiologic factors contributing to the diversities of the human host. In a broad sense, in epidemiology, the classification of human populations is normally based on gender, race, and ethnicity. The term race describes populations or groups of people as distinguished by various sets of characteristics and beliefs about common ancestry. The most widely used human racial categories are based on visible traits (especially skin color, facial features and hair texture), and self-identification *(67)*. The term ethnicity refers to a population of human beings whose members identify with each other, either on the basis of a presumed common genealogy or ancestry, or recognition by others as a distinct group, or by common cultural, linguistic, religious, or territorial traits. Members of an ethnic group, on the whole, claim cultural continuity over time and this continuity has been a prime tool and of great interest to epidemiologists studying the variation of disease rates among populations when they mix and migrate *(68)*. The diversity in cancer rates based on race and ethnicity is well studied, especially in countries such as the United States, with a great deal of racial diversity and qualified cancer data. Here, the disparities of public health indicators among the black and white populations are almost universal and the same is true for cancer, from different patterns of cancer occurrence to different use patterns of cancer screening, as well as receiving varying qualities of treatment. Based on available data, cancer incidence among black males exceeds that of whites for about half of the individual cancers registered, with the combined rate being 20% higher *(69)*. Esophageal cancer has the largest relative difference, three times higher for black males than whites *(11)*. Alcohol and tobacco use are the prime risk factors for esophageal cancer, both of which are higher among blacks than whites *(53)*. There are thorough studies of cancer occurrence among Ashkenazi Jews, with the understanding that

the genetic ingredient plays a major role in the higher incidence of early onset breast cancer in this population *(70)*. The variation in cancer incidence seen among different ethnic populations has been studied at the local and international levels to provide answers to very basic questions regarding the degree to which cancer risk is attributed to the genetic ingredient versus external factors including lifestyle and environment.

There are basically two major lines of studies addressing the effect of genes, environmental factors, or both in the etiology of cancer. The first line deals with studies of human population genetics by identifying genes or traits associated with certain cancers. These studies normally use linkage and segregation methodologies to address genes in relation to disease occurrence. The second line of studies that have targeted variations in cancer rates among human populations are basically twin, adoption, and, importantly, migrant studies. In both lines of studies, the main aim is to address the disparities of diseases or cancer occurrence among the human population to develop hypotheses and understand the etiologic factors responsible for the occurrence of cancer or contributing to cancer control and treatment. Although the first line contributes a great deal to the identification of genes in the etiology of cancer, the second line of human population studies are more involved with the external and modifiable component of cancer etiology. Some public health experts consider genes as a nonmodifiable ingredient in the etiology of any disease, including cancer, and they emphasize external factors or the interaction of external factors with the nonmodifiable genetic ingredient (known as the gene–environment interaction) in addressing cancer control strategies. Because the first line has more of a human genetic orientation, this chapter deals more with the second line, especially the migrant studies, to expose the reader to the potentials of these kinds of studies in addressing the etiology of cancer.

4.1. Migrant Studies

When individuals migrate from one region to another, the rate of cancer occurrence among them changes from that of their country of origin, and it approaches that of their new host country *(71)*. Studies of these types of changes in incidence rate are called migrant studies. The rationale underlying migrant studies in cancer is to compare the frequency of cancer between members of the same ethnic group (who are assumed to be genetically homogeneous) who are living in different environments. The migration of a population to a new country and the rapid changes in their exposure to new hazards bring very unique opportunities, first, to study the effects of heredity and environment that are clues regarding the natural history of an exposure and outcome; second, to identify host- and environment-related protective and risk factors; and third, to identify new

subpopulations of migrants that may differ in some background host-related factors due to new exposures. In addition, migrant population studies have been used to validate reported rates or even estimate cancer rates in the country of origin, because most immigration is basically migration from low-resource countries (where an accurate rate of cancer may not be available) to affluent countries where organized registries of health events (including cancer) exist. Different methodologies, used to study the changes of the cancer occurrence in immigrant populations, include the following:

1. Studies of cancer mortality and morbidity that compare rates in the
   - immigrant population with the population in the country of adaptation,
   - immigrant population with the population in their country of origin, and
   - first generation of the immigrant population with the second generation.
2. Comparisons of immigrant to nonimmigrant populations for a known exposure to carcinogens.
3. Studies of genetic traits and polymorphic genes in the immigrant population and their comparison with the host or adaptive country (impressibility of the trait and gene–environment interaction).
4. Studies of cancer precursors in the immigrant population.
5. Studies comparing cancer rates between different groups of immigrants from different countries of origin.

The main and major migrant studies have been the studies of mortality and incidence among migrants to their country of adaptation. Almost all cancer sites have been studied among immigrants. The most frequently studied cancer site is the breast. Most migrant incidence studies of breast cancer involved populations that have migrated from low-risk to high-risk countries. Breast cancer incidence is very low among Japanese, and studies involving Japanese who have immigrated to Hawaii and California have shown a two-fold increase in rates among the first generation of immigrants compared with their counterparts in the existing population during the same time *(72,73)*. The rate of breast cancer has increased in the same immigrant population among the second generation for all ages *(72,73)*. Additionally, the same pattern has been seen in the Chinese *(74)* and Spanish *(75)* who have migrated to United States, and in Asians who have migrated to Israel *(76)*. The changes in incidence among white Caucasians who have immigrated to the United States are different from those of other skin colors; the rate of Caucasians from England and Wales who immigrated to the United States has

sharply changed to match that of American Caucasians. Although there are few studies of migration from high-risk to low-risk areas, one such study of migrants from the high-risk countries of England, Wales, and Ireland to low-risk Australia has shown that the migrant women experience a higher rate than the native-born Australians *(77)*.

Not all the migrant studies have shown changes in rate toward that of the country of adaptation *(72)*. The migrant studies of breast cancer indicate that exposure to a new environment as an adult can influence your risk of breast cancer. The varying degrees of rate changes among white versus colored skin may indicate both genetic and environmental factors (differential levels of exposure due to varying lifestyles between different migrant groups). The risk of breast cancer among the immigrants depends on the age at migration: the younger the immigrant the higher the risk of breast cancer, proving a dose–response relationship between environment and risk of breast cancer *(76)*.

The studies of cancer mortality among the immigrant population are of great interest, and the interpretation of mortality data requires special consideration. The study of mortality from breast cancer among immigrants has shown a different pattern from that of incidence *(78)*; breast cancer mortality rates in Polish immigrants in the United States, England, Wales *(79)*, and Australia *(80)* were higher than those in Poland and similar to those of the native-born population, whereas Polish immigrants had intermediate rates between the home and host countries in France *(81)*. The mortality studies among immigrants have inherent problems originating from mortality data, especially that mortality is influenced by the level of health care use, screening practices, and the validity of establishing cause of death.

4.1.1. Methodologic Consideration of Migrant Studies

All available epidemiologic methodologies have been used in the study of migrant populations. Although the incidence and mortality studies have been correlational (i.e., comparing aggregate data or rates), they have provided clues to the very simple question, How important is the role of genetic factors versus environmental factors in the etiology of cancer? Migrant studies based on individual cases, such as cohorts or case controls, contain information about known risk factors that are ascertained at the individual level, and more specific questions, such as those regarding the role of certain lifestyles and behaviors or nutritional habits and intake, can be studied with no risk of ecological fallacies. The migrant study can be considered a natural experiment in terms of its power in avoiding bias and its efficiency in answering complicated epidemiologic questions about the etiology of cancer. Migrant studies can provide more extreme variations of exposure in two groups of genetically similar subjects, thereby minimizing the chances of misclassification of exposure (in the case where

migrants are compared with their descendants) and selection bias (as the migrant studies are structurally cohort-based studies and less subject to selection bias). When the aggregated data of migrant populations are analyzed, the uncertainties that arise are those of correlational studies and can easily be subject to ecological fallacy, especially when the third and fourth generation of naturalized immigrants are studied using the aggregated data.

Other methodological issues in migrant studies are how to ascertain the migrant status of people. Several different methods have been used for such ascertainment, each with its own validity and limitation; when aggregated information is analyzed, normally the number of patients comes from the registries and the population comes from the databases of the immigration office. The immigration offices, especially in developed countries, keep demographic information about individual immigrants, including name, surname, city and country of birth, ethnicity, year of immigration and age at immigration, plus some other socioeconomic indicators. This information is linked to cancer registry databases to identify cancer cases among the immigrant population. Although migrant studies have used immigration office data banks to ascertain rates or immigration status, other means of establishing immigration status have been used with different degrees of sensitivity and specificity. One example involves the ascertainment of immigrant status based on such loose definitions such as having a Latino surname *(82)*. The method of ascertaining immigration status, the availability of detailed information about individual immigrant and the validity of this information are the issues pertaining to the methodological aspect of the migrant studies *(83,84)*.

4.1.2. Interpretation of Migrant Studies

Migrant studies are of great value in the epidemiologic investigation of the etiology of cancer because migrants provide unique research opportunities that can rarely be achieved in the absence of migration. In the study of immigrants, when the aggregated data of incidence and mortality are studied, the changes in the rate of cancer incidence or mortality fall under three scenarios. The first scenario involves a cancer rate increase, as has been the case for many cancer sites, among the immigrant population compared with that of the population from their homeland who did not migrate. In this scenario, three possible interpretations include a new exposure, a higher level of exposure, or the higher susceptibility to the exposure affecting the immigrant population. The incidence study alone may not be able to attribute the increase to a special exposure unless other individual-based studies, such as cohort or case-control studies, are undertaken. The second scenario involves a rate decrease among the immigrant population compared with the host population. This happens when the people in the country of origin experience a higher

rate of certain cancer compared with the host countries. In interpreting this rate decrease, one needs to exclude the possibility of treatment of cancer precursors detected early in the host country, for such early detection may not be possible in the immigrants native country. When there is a genuine decrease in the rate of cancer among immigrants, the second generation should experience another decrease compared to the first generation, if the external and environmental factors are to be attributed to changes. A genuine decrease implies the existence of a protective factor or lack of a risk in the host country. As mentioned in the first scenario, a residual maintenance of the lower rate among the third generation may indicate a genetic gradient for the cancer. The third scenario occurs when the rate of cancer does not change, this scenario implies that either there is no difference in the exposure level and circumstances between the two countries or the cancer of interest has a strong genetic gradient that acts independently of external exposure.

In interpreting rate changes arising in each scenario, one must attribute the changes to a factor. Although several categories of risks or factors may influence the changes in rate within the migrated population, the diet and nutritional habits (the major causally attributed factor after smoking) are strong confounders in migrant studies, as there is a marked change in the availability of nutrients in the host country plus the immigrant population cannot keep their original nutritional habit in the host country. In the design and interpretation of migrant studies, especially when the immigrant populations are studied at the individual level, special attention should be given to diet and nutritional habits. Another problem that must be carefully considered when rate changes are seen among an immigrant population is that the host country and native population may differ in the rate of screening, and a lead time bias may contribute to the increasing prevalence and incidence of cancer cases detected early.

An increase in cancer rate from the first to the second generation is stronger evidence of the effect of new exposure in the host country and the smaller influence of genetic factors in the etiology of the studied cancer. Although studies of the third generation are difficult as the migrated population mixes with the host population, any differences in the magnitude of rate between the migrated population and the host population may indicate a genetic gradient of the studied cancer.

Another issue that may introduce uncertainty in interpreting migrant studies is the effect of immigration regulations on the variation seen among immigrants; such variation may create subgroups in terms of socioeconomic or other factors related to the outcome of the study or cancer. Such subgroups, if not detected and ascertained in the study, may introduce uncertainty and bias in the study and complicate the interpretation. Another issue

that needs to be considered regarding migrant studies is that the immigrant population may come from a special class of a society and in fact they may differ in their rate of cancer within their own homeland population. Such a phenomenon is analogous to the healthy worker effect and it can be called a "migrant effect."

The last problem to be considered in migrant studies is that, when comparing the immigrant rate of cancer in the host country to their rate in the homeland, the rates in the two populations may be estimated based on different methodologies with regard to case ascertainment, diagnostic criteria, screening practice, validity of cause of death, classification comparability, and other country specific issues. Such discrepancies have always been challenging issues in migrant studies.

Another important issue in the interpretation of cancer rate among the migrant is that migrant studies must be done for site-specific cancers and performing studies based on all cancers combined may introduce serious bias as some cancers may be decreasing and some increasing among migrants, which will dilute the changes that can otherwise be seen with site-specific cancers.

References

1. Johnson S (2006) The Ghost Map: The Story of London's Most Terrifying Epidemic–and How it Changed Science, Cities and the Modern World. Riverhead Books, New York, NY
2. Parkin DM (2006) The evolution of the population-based cancer registry. *Nat. Rev. Cancer* **6,** 603–612.
3. Jensen OM, Parkin DM, Maclennan R., Muir C.S., Skeet R.G. (1991) Cancer Registration Principals and Methods. International Agency for Research on Cancer, Lyon, France.
4. Weeber M, Kors JA, Mons B (2005) Online tools to support literature-based discovery in the life sciences. *Brief. Bioinformatics* **6,** 277–286.
5. Srinivasan P, Libbus B (2004) Mining MEDLINE for implicit links between dietary substances and diseases. *Bioinformatics* **20(Suppl 1),** i290–i296.
6. Rebholz-Schuhman D, Cameron G, Clark D, van ME, Coatrieux JL, Del Hoyo BE, Martin-Sanchez F, Milanesi L, Porro I, Beltrame F, Tollis I, Van der LJ (2007) SYMBIOmatics: synergies in medical informatics and bioinformatics—exploring current scientific literature for emerging topics. *BMC Bioinformatics* **8(Suppl 1),** S18.
7. Schwartz J (1990) Multinational trends in multiple myeloma. *Ann. N Y Acad. Sci.* **609,** 215–224.
8. Davis DL, Hoel D, Percy C, Ahlbom A, Schwartz J (1990) Is brain cancer mortality increasing in industrial countries? *Ann. N Y Acad. Sci.* **609,** 191–204.
9. Gonzalez-Diego P, Lopez-Abente G, Pollan M, Ruiz M (2000) Time trends in ovarian cancer mortality in Europe (1955–1993): effect of age, birth cohort and period of death. *Eur. J. Cancer* **36,** 1816–1824.
10. Aragones N, Pollan M, Rodero I, Lopez-Abente G (1997) Gastric cancer in the European Union (1968–1992): mortality trends and cohort effect. *Ann. Epidemiol.* **7,** 294–303.
11. Jemal A, Siegel R, Ward E, Murray T, Xu J, Smigal C, Thun MJ (2006) Cancer statistics, 2006. *CA Cancer J. Clin.* **56,** 106–130.
12. Howe HL, Wu X, Ries LA, Cokkinides V, Ahmed F, Jemal A, Miller B, Williams M, Ward E, Wingo PA, Ramirez A, Edwards BK (2006) Annual report to the nation on the status of cancer, 1975–2003, featuring cancer among U.S. Hispanic/Latino populations. *Cancer* **107,** 1711–1742.
13. Thun MJ, Jemal A (2006) How much of the decrease in cancer death rates in the United

States is attributable to reductions in tobacco smoking? *Tob. Control* **15,** 345–347.

14. Althuis MD, Dozier JM, Anderson WF, Devesa SS, Brinton LA (2005) Global trends in breast cancer incidence and mortality 1973–1997. *Int. J. Epidemiol.* **34,** 405–412.
15. Lacey JV Jr, Devesa SS, Brinton LA (2002) Recent trends in breast cancer incidence and mortality. *Environ. Mol. Mutagen.* **39,** 82–88.
16. Steliarova-Foucher E, Stiller C, Kaatsch P, Berrino F, Coebergh JW (2005) Trends in childhood cancer incidence in Europe, 1970–99. *Lancet* **365,** 2088.
17. Steliarova-Foucher E, Stiller C, Kaatsch P, Berrino F, Coebergh JW, Lacour B, Parkin M (2004) Geographical patterns and time trends of cancer incidence and survival among children and adolescents in Europe since the 1970s (the ACCISproject): an epidemiological study. *Lancet* **364,** 2097–2105.
18. Mosavi-Jarrahi A, Mohagheghi MA (2006) Epidemiology of esophageal cancer in the high-risk population of Iran. *Asian Pac. J. Cancer Prev.* **7,** 375–380.
19. Parkin DM (2006) The evolution of the population-based cancer registry. *Nat. Rev. Cancer* **6,** 603–612.
20. Ravakhah K (2006) Death certificates are not reliable: revivification of the autopsy. *South. Med. J.* **99,** 728–733.
21. Modelmog D, Rahlenbeck S, Trichopoulos D (1992) Accuracy of death certificates: a population-based, complete-coverage, one-year autopsy study in East Germany. *Cancer Causes Control* **3,** 541–546.
22. Hasuo Y, Ueda K, Kiyohara Y, Wada J, Kawano H, Kato I, Yanai T, Fujii I, Omae T, Fujishima M (1989) Accuracy of diagnosis on death certificates for underlying causes of death in a long-term autopsy-based population study in Hisayama, Japan; with special reference to cardiovascular diseases. *J. Clin. Epidemiol.* **42,** 577–584.
23. Kircher T, Nelson J, Burdo H (1985) The autopsy as a measure of accuracy of the death certificate. *N. Engl. J. Med.* **313,** 1263–1269.
24. Schelhase T, Weber S (2007) [Mortality statistics in Germany. Problems and perspectives.]. *Bundesgesundheitsblatt. Gesundheitsforschung. Gesundheitsschutz* **50,** 969–976.
25. Jougla E, Pavillon G, Rossollin F, De SM, Bonte J (1998) Improvement of the quality and comparability of causes-of-death statistics inside the European Community. EUROSTAT Task Force on "causes of death statistics". *Rev. Epidemiol. Sante Publique* **46,** 447–456.
26. Cano-Serral G, Perez G, Borrell C (2006) Comparability between ICD-9 and ICD-10 for the leading causes of death in Spain. *Rev. Epidemiol. Sante Publique* **54,** 355–365.
27. Anderson RN, Minino AM, Hoyert DL, Rosenberg HM (2001) Comparability of cause of death between ICD-9 and ICD-10: preliminary estimates. *Natl. Vital Stat. Rep.* **49,** 1–32.
28. Chandramohan D, Maude GH, Rodrigues LC, Hayes RJ (1994) Verbal autopsies for adult deaths: issues in their development and validation. *Int. J. Epidemiol.* **23,** 213–222.
29. Setel PW, Sankoh O, Rao C, Velkoff VA, Mathers C, Gonghuan Y, Hemed Y, Jha P, Lopez AD (2005) Sample registration of vital events with verbal autopsy: a renewed commitment to measuring and monitoring vital statistics. *Bull. World Health Organ.* **83,** 611–617.
30. Owusu-Agyei S, Awini E, Anto F, Mensah-Afful T, Adjuik M, Hodgson A, Afari E, Binka F (2007) Assessing malaria control in the Kassena-Nankana district of northern Ghana through repeated surveys using the RBM tools. *Malar. J.* **6,** 103.
31. Verdecchia A, Capocaccia R, Egidi V, Golini A (1989) A method for the estimation of chronic disease morbidity and trends from mortality data. *Stat. Med.* **8,** 201–216.
32. Capocaccia R, Verdecchia A, Micheli A, Sant M, Gatta G, Berrino F (1990) Breast cancer incidence and prevalence estimated from survival and mortality. *Cancer Causes Control* **1,** 23–29.
33. Micheli A, Verdecchia A, Capocaccia R, De AG, Gatta G, Sant M, Valente F, Berrino F (1992) Estimated incidence and prevalence of female breast cancer in Italian regions. *Tumori* **78,** 13–21.
34. Kerlikowske K, Miglioretti DL, Buist DS, Walker R, Carney PA (2007) Declines in invasive breast cancer and use of postmenopausal hormone therapy in a screening mammography population. *J. Natl. Cancer Inst.* **99,** 1335–1339.
35. Minelli L, Stracci F, Prandini S, Moffa IF, La RF (2004) Gynaecological cancers in Umbria (Italy): trends of incidence, mortality and survival, 1978–1998. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **115,** 59–65.
36. Cox B, Skegg DC (1992) Projections of cervical cancer mortality and incidence in

New Zealand: the possible impact of screening. *J. Epidemiol. Community Health* **46,** 373–377.

37. Robertson C, Boyle P (1998) Age-period-cohort analysis of chronic disease rates. I: Modelling approach. *Stat. Med.* **17,** 1305–1323.
38. Robertson C, Boyle P (1998) Age-period-cohort models of chronic disease rates. II: Graphical approaches. *Stat. Med.* **17,** 1325–1339.
39. Hristova L, Hakama M (1997) Effect of screening for cancer in the Nordic countries on deaths, cost and quality of life up to the year 2017. *Acta Oncol.* **36(Suppl 9),** 1–60.
40. Miettinen O, Wang J (1981) An alternative to the proportionate mortality ratio. *Am J Epidemiology* **114,** 144–148.
41. Yazdizadeh B, Jarrahi AM, Mortazavi H, Mohagheghi MA, Tahmasebi S, Nahvijo A (2005) Time trends in the occurrence of major GI cancers in Iran. *Asian Pac. J. Cancer Prev.* **6,** 130–134.
42. Ward EM, Thun MJ, Hannan LM, Jemal A (2006) Interpreting cancer trends. *Ann. N Y Acad. Sci.* **1076,** 29–53.
43. Ajiki W, Tsukuma H, Oshima A (1998) [Index for evaluating completeness of registration in population-based cancer registries and estimation of registration rate at the Osaka Cancer Registry between 1966 and 1992 using this index]. *Nippon Koshu Eisei Zasshi* **45,** 1011–1017.
44. Parkin DM (2006) The evolution of the population-based cancer registry. *Nat. Rev. Cancer* **6,** 603–612.
45. Gonzalez-Diego P, Lopez-Abente G, Pollan M, Ruiz M (2000) Time trends in ovarian cancer mortality in Europe (1955–1993): effect of age, birth cohort and period of death. *Eur. J. Cancer* **36,** 1816–1824.
46. Wu K, Li K (2007) Association between esophageal cancer and drought in China by using Geographic Information System. *Environ. Int.* **33,** 603–608.
47. Garb JL, Ganai S, Skinner R, Boyd CS, Wait RB (2007) Using GIS for spatial analysis of rectal lesions in the human body. *Int. J. Health Geogr.* **6,** 11.
48. Bunnell JE, Tatu CA, Bushon RN, Stoeckel DM, Brady AM, Beck M, Lerch HE, McGee B, Hanson BC, Shi R, Orem WH (2006) Possible linkages between lignite aquifers, pathogenic microbes, and renal pelvic cancer in northwestern Louisiana, USA. *Environ. Geochem. Health* **28,** 577–587.
49. Thompson JA, Carozza SE, Zhu L (2007) An evaluation of spatial and multivariate covariance among childhood cancer histotypes in Texas (United States). *Cancer Causes Control* **18,** 105–113.
50. Briggs DJ, Elliott P (1995) The use of geographical information systems in studies on environment and health. *World Health Stat. Q.* **48,** 85–94.
51. Nasca C.Philip, Pastides Harris (2001) Fundamentals of Cancer Epidemiology. Aspen Publishers, Inc., Gaithersburg, MD.
52. Muir C.S., Nectoux J. (1996) International patterns of cancer. *In* Cancer Epidemiology and Prevention, Schottenfeeld D and Fraumeni JF (eds.). Oxford University Press, Oxford, NY, pp. 141–167.
53. Munoz N., Day N.E. (1996) Esophageal cancer. *In* Cancer Epidemiology and Prevention Schottenfeeld D and Fraumeni JF (eds.). Oxford University Press, Oxford, NY, pp. 681–685.
54. Beck L.R., Kitron U., Bobo M. (2002) Remote sensing, GIS, and spatial statistics: powerful tools for landscape epidemiology. *In* Health Impacts of Global Environmental Change: Concepts and Methods, Martens P (ed.). Cambridge University Press, Cambridge, New York, NY.
55. Parkin DM, Whelan SL, Ferlay J, Teppo L, and Thomas DB (2002). Cancer incidence in five continents. [VIII]. IARC Scientific Publications No. 155. IARC Press, Lyon, France.
56. Swerdlow A, Dos-Santos SI (1993) Atlas of Cancer Incidence in England and Wales1968–85. Oxford University Press, Oxford, New York.
57. Walter SD, Birnie SE (1991) Mapping mortality and morbidity patterns: an international comparison. *Int. J. Epidemiol.* **20,** 678–689.
58. Jarup L (2004) Health and environment information systems for exposure and disease mapping, and risk assessment. *Environ. Health Perspect.* **112,** 995–997.
59. Walter SD (1993) Visual and statistical assessment of spatial clustering in mapped data. *Stat. Med.* **12,** 1275–1291.
60. Clarke KC, McLafferty SL, Tempalski BJ (1996) On epidemiology and geographic information systems: a review and discussion of future directions. *Emerg. Infect. Dis.* **2,** 85–92.
61. Gatrell AC, Bailey TC (1996) Interactive spatial data analysis in medical geography. *Soc. Sci. Med.* **42,** 843–855.
62. Jacquez GM (2004) Current practices in the spatial analysis of cancer: flies in the ointment. *Int. J. Health Geogr.* **3,** 22.

63. Alexander FE, Boyle P, Carli PM, Coebergh JW, Draper GJ, Ekbom A, Levi F, McKinney PA, McWhirter W, Michaelis J, Peris-Bonet R, Petridou E, Pompe-Kirn V, Plisko I, Pukkala E, Rahu M, Storm H, Terracini B, Vatten L, Wray N (1998) Spatial clustering of childhood leukaemia: summary results from the EUROCLUS project. *Br. J. Cancer* **77,** 818–824.
64. Little J (1999) Epidemiology of Childhood Cancer. IARC Press, Lyon, France.
65. Kulldorff M (2004) SaTScan 5.1. Information Management Services, Inc., Boston, MA .
66. Mosavi-Jarrahi A, Moini M, Mohagheghi MA, Alebouyeh M, Yazdizadeh B, Shahabian A, Nahvijo A, Alizadeh R (2007) Clustering of childhood cancer in the inner city of Tehran metropolitan area: a GIS-based analysis. *Int. J. Hyg. Environ. Health* **210,** 113–119.
67. Keita SO, Kittles RA, Royal CD, Bonney GE, Furbert-Harris P, Dunston GM, Rotimi CN (2004) Conceptualizing human variation. *Nat. Genet.* **36,** S17–S20.
68. Rivara F, Finberg L (2001) Use of the terms race and ethnicity. *Arch. Pediatr. Adolesc. Med.* **155,** 119.
69. Horn JW, Devesa SS, Burhansstipanov L (1996) Cancer Incidence, Mortality, and Survival among Racial and Ethnic Minority groups in the United States. *In* Cancer Epidemiology and Prevention, Schottenfeeld D and Fraumeni JF (eds.). Oxford University Press, Oxford, New York.
70. Rubinstein WS (2004) Hereditary breast cancer in Jews. *Fam. Cancer* **3,** 249–257.
71. Thomas D, Karagas M (1996) Migrant studies. *In* Cancer Epidemiology and Prevention. Oxford University Press, Oxford, New York, pp. 236–266.
72. Kolonel LN (1980) Cancer patterns of four ethnic groups in Hawaii. *J. Natl. Cancer Inst.* **65,** 1127–1139.
73. Buell P (1973) Changing incidence of breast cancer in Japanese-American women. *J. Natl. Cancer Inst.* **51,** 1479–1483.
74. Stanford JL, Herrinton LJ, Schwartz SM, Weiss NS (1995) Breast cancer incidence in Asian migrants to the United States and their descendants. *Epidemiology* **6,** 181–183.
75. De SE, Parkin DM, Khlat M, Vassallo A, Abella M (1990) Cancer in migrants to Uruguay. *Int. J. Cancer* **46,** 233–237.
76. Parkin DM, Steinitz R, Khlat M, Kaldor J, Katz L, Young J (1990) Cancer in Jewish migrants to Israel. *Int. J. Cancer* **45,** 614–621.
77. McMichael AJ, Giles GG (1988) Cancer in migrants to Australia: extending the descriptive epidemiological data. *Cancer Res.* **48,** 751–756.
78. Staszewski J, Haenszel W (1965) Cancer mortality among the Polish-born in the United States. *J. Natl. Cancer Inst.* **35,** 291–297.
79. Adelstein AM, Staszewski J, Muir CS (1979) Cancer mortality in 1970–1972 among Polish-born migrants to England and Wales. *Br. J. Cancer* **40,** 464–475.
80. Staszewski J, McCall MG, Stenhouse NS (1971) Cancer mortality in 1962–66 among Polish migrants to Australia. *Br. J. Cancer* **25,** 599–610.
81. Tyczynski J, Parkin D, Zatonski W, Tarkowski W (1992) Cancer mortality among Polish migrants to France. *Bull. Cancer* **79,** 789–800.
82. Stewart SL, Swallen KC, Glaser SL, Horn-Ross PL, West DW (1999) Comparison of methods for classifying Hispanic ethnicity in a population-based cancer registry. *Am. J. Epidemiol.* **149,** 1063–1071.
83. Swallen KC, West DW, Stewart SL, Glaser SL, Horn-Ross PL (1997) Predictors of misclassification of Hispanic ethnicity in a population-based cancer registry. *Ann. Epidemiol.* **7,** 200–206.
84. Swallen KC, Glaser SL, Stewart SL, West DW, Jenkins CN, McPhee SJ (1998) Accuracy of racial classification of Vietnamese patients in a population-based cancer registry. *Ethn. Dis.* **8,** 218–227.

Chapter 8

Evaluation of Environmental and Personal Susceptibility Characteristics That Modify Genetic Risks

Jing Shen

Summary

Study design in understanding gene–environment interaction plays a crucial role. Different study designs with their advantages and limitations are described in this chapter. Gene penetrance has been studied in several cancers, including breast and prostate cancer. Compared with high-penetrance genes, such as breast cancer (BRCA)1, BRCA2 in breast cancer, gene–environment interaction plays a major role in cancer development where low-penetrance genes are the major players. Genetic polymorphism is determined in low- and high-penetrance genes to identify cancer-related polymorphisms. The role of genetic and epigenetic factors in cancer development is discussed. Preventive approaches, especially in the epigenetic field, show promise. A discussion about different epidemiological methods with examples is provided.

Key Words: Biomarker, classification and regression tree, environmental factors, epigenetics, genetic risk, high-penetrance genes, personal susceptibility.

1. Introduction

Most complex diseases are the result of the interactions between multiple genes (genetic agents) and environmental factors (endogenous and exogenous exposures) and personal susceptibility characteristics (personal lifestyle and behavior). Cancer, as one of human complex diseases, also is determined by these factors and their interactions *(1)*. Although a great deal of interest has been devoted to explore human "major cancer genes" with a high penetrance, such as Breast cancer (*BRCA)1* or *BRCA2* for hereditary breast cancer, and adenomatous polyposis coli, *MLH1*, or *MSH2* for colorectal cancer, the diffusion of such genetic mutations in the population is rare, and their impact on

Mukesh Verma (ed.), *Methods in Molecular Biology, Cancer Epidemiology, vol. 471*
© 2009 Humana Press, a part of Springer Science + Business Media, Totowa, NJ
Book doi: 10.1007/978-1-59745-416-2

the general cancer risk, from a public health perspective, is limited. It is estimated that <5% of cancer is attributable to these high-penetrance genes. Therefore, >95% sporadic cancer should be the major concern of cancer epidemiologic studies because it consists of most of the cancer cases in population.

The results obtained from epidemiologic case-control study, cohort study, twin or family study, and migrant study over one or two generations have demonstrated the unquestionable role of environmental factors (e.g., personal lifestyle and behavior factors) in tumorigenesis of sporadic cancer, and 75–90% of sporadic cancer has been thought to be environmentally determined. In epidemiologic terms, the population attributable risk (i.e., the proportion of cancer in the population) of environmental factors is considered to be up to 90% for all types of cancer. Despite considerable effort, scientists have not been able to identify common highly penetrant genes in sporadic cancer (maybe it does not exist fundamentally). The failure to provide etiologic evidence for genetic factors in sporadic cancers, even in the well-designed studies, simply indicates the low heritability of these cancers under study. Recently, a relatively new field of research has focused on more common genetic alterations that by themselves may not substantially impact cancer risk but that in concert with environmental exposures or personal characteristics may lead to development of cancer. The most common form of the genetic alteration is polymorphisms. Genetic polymorphisms have resulted from genetic mutations that have survived and passed through generations, and which have a prevalence of at least 2% in a population. However, the independent role of these genetic factors and their interaction with environmental factors in the occurrence of sporadic cancers is still under study *(2)*.

Gene and environment are biologically inseparable, the impact of a person's environment is translated through the genome, and the impact of the genome on cancer risk depends on the environment. Recently, the epidemiologic description for human cancer causation has become more complicated along with the intensive studies. First, it has been shown that some known environmental risk factors, such as tobacco smoking and asbestos exposure, interact and jointly create a higher risk of lung cancer than the sum of the separate risks. Second, a uniform exposure would elicit different effects depending on an individual's genetic background. The effects of environmental exposures might be transmitted at the cellular level through mechanisms that could vary among individuals with different genetic and epigenetic status. There is also interaction between the gene product(s), the by-products of the environmental insult, or a combination to influence cancer risk. Although the extent of the genetic involvement and their interaction with environment in tumorigenesis is still elusive, many known carcinogens could potentially elicit heritable

effects at many levels, including carcinogen metabolism, DNA repair, cell cycle control, apoptosis, genomic instability, and epigenetic changes. With the development of molecular techniques and statistical skills, comprehensive evaluations for the influence of genetic and environmental factors, and their multiple interactions on cancer risk are available. In this chapter, I introduce and discuss some aspects of evaluating the role of environmental and personal susceptibility characteristics in modifying genetic risks for sporadic cancers.

2. Genetic and Environmental Factors for Cancer Risk

In classic epidemiologic studies, the environmental risk factors for cancer (carcinogens) are known to include factors in physical, chemical, and biological aspects obtained mostly from questionnaires, which are often used with a broad scope in the medical literature, such as diet, lifestyle, and infectious agents. In this broad sense, the environment is implicated in the causation of the majority of human cancers. Although several well-documented environmental variables have been confirmed to influence cancer risk, such as tobacco smoking and lung cancer, hepatitis B virus (HBV) infection, hepatocellular carcinoma, and gastric cancer because of bacterium *Helicobacter pylori*, quantifying these and most of other environmental factors with precision is difficult. Most of the environmental factors assessed by questionnaire may be underreported, e.g. smoking or alcohol consumption, and overreported in the case of exercise. In addition, it is hard to adjust for factors such as the variable tar content of cigarettes, which might alter the impact of smoking. The most efficient and commonly used tool to collect data of food and nutrient intake at the population level is the food frequency questionnaire, but assessing the fat content and fatty acid composition of food is not straightforward. Besides the environmental factors, genetic factors observed in classic epidemiologic studies also are obtained from questionnaire data. This traditional approach restricted the evaluation of genetic factors to ethnicity, gender, family history of cancers in first or second degree relatives, and characteristics of twins (monozygotic or dizygotic twin).

The rapidly developed molecular epidemiologic approaches, in recent decades, expanded the field to allow more accurate exposure assessment, improved understanding of genetic susceptibility and intermediate endpoints, and enhanced risk prediction by using both genetic and environmental biomarkers, although classic epidemiologic studies also involved in a few. The introduction of biological markers has been perceived as a great improvement

in terms of understanding the biological mechanisms behind an association between a risk factor and cancer. Biomarkers used for cancer molecular epidemiologic research can be categorized into several classes: exposure biomarkers (HBV, hepatitis C virus, and *H. pylori* infection), biomarkers of internal dose (DNA adducts formed after the metabolic activation of certain carcinogens such as aflatoxin B1), biomarkers of biologically effective dose (somatic *p53* mutations that may both indicate exposure to a specific carcinogen and be one of the genetic "hits" in the multihit model of carcinogenesis, polycyclic aromatic hydrocarbons, PAH-DNA adducts), phenotype biomarkers (mutagen sensitivity, comet assay, host cell reactivation assay, and so on, for the phenotype of DNA repair capacity), biomarkers of altered structure/function (chromosomal aberrations and sister chromatid exchanges), biomarkers of susceptibility (metabolic polymorphisms in genes that are involved in carcinogen metabolism or detoxification, polymorphism in DNA repair pathway), biomarker of genomic instability (telomeres length and telomerase activity), biomarkers of cancer subtypes (estrogen and progesterone receptors in breast cancer), biomarkers of prognosis (metabolic polymorphisms in drug-metabolism gene), and epigenetic biomarkers (DNA methylation and histone modification).

In practice, we must notice the limitations of application of some biomarkers in epidemiologic studies. In case-control studies, biomarkers of exposure, effect, or both could be found in cases as a consequence of cancer, or the therapy; therefore, they may not be a sign or consequence of the risk factor/exposure of interest. In addition, biomarkers reflect changes that have happened in the recent past, a time that is not relevant for the development of cancer, where the length of intervals between exposure and cancer is usually measurable in a decade or more. In cohort studies, a one-time measure that is performed at entry into the study may not be representative of the subject's chronic exposure to a certain risk factor. Some genetic or epigenetic biomarkers (methylation status, telomere dysfunction) also may be influenced by environmental exposure, and they are changeable during the lifetime, although related data are very limited. However, most of genetic markers, especially the germline mutations or the polymorphisms, are present at birth. The measurement can be performed at any time during the whole life. Most previous cancer molecular epidemiology studies have used biomarkers to study a single or at most only a few genetic or environmental factors for sporadic cancer. However, tumorigenesis is a multifactorial and multistage process, which is dependent on multiple mechanisms and pathways that regulate absorption, metabolic activation, and excretion, DNA repair, control of cell cycle and inflammatory responses to carcinogens. Genetic variability in any of these pathways can alter an individual's inherited predisposition to

environmental exposure. The inability to replicate previous results is another common problem in this kind of association study. Explanations include false-positive results that have occurred by chance, inadequate statistical power, and true variation in underlying the association among different populations. Therefore, in cancer risk assessment, consideration of any single biomarker is not sufficient to capture the full extent of the mediating effects of susceptibility genes.

3. Gene–Environment Interaction for Cancer Risk

Classic epidemiologic studies have used a "black box" approach to sort out cancer risks that makes it difficult to identify gene–environment interactions in cancer causation. The advent of molecular epidemiology incorporating tools from molecular biology, genetics, cytogenetics, biostatistics, and bioinformatics has allowed scientists to effectively evaluate gene–environment interactions *(3)*. Although assessment of gene–environment interaction effects without estimating and understanding the main effects of the genetic and environmental factors will be of little use for public health or individual risk assessment, there is still potential significance for increasing knowledge about biologic mechanisms underlying gene–environment interaction.

Interaction is a model-dependent concept. One working model may be helpful to better understand the complicated interactions of genetic and environmental factors (**Fig. 8.1**). The hypothesis is that there is a range of genetic risk profiles in the population, with each individual occupying a position along the risk spectrum (from a low to a high genetic risk). Similarly, individuals adopt a different position on the environmental spectrum of risk by the environment they live and lifestyle choices they make. The importance of gene–environment interaction is that only when an individual with a high-risk genetic profile enters a high-risk environment, will the effect on risk be so great that cancer develops.

In a case-controls study, the gene–environment interaction may be assessed on a multiplicative scale by calculating odds ratios (Ors) for genotypes stratified by indicators of exposures (such as cigarette smoking, intake of antioxidants levels or supplemental antioxidants), running separate models by including multiplicative interaction terms in the logistic regression model. If this term is significant, then this implies statistically significant interaction. Evaluation of interactions on an additive scale between genotypes and exposures will be performed by using indicator terms for those with the genotype only, exposure only, and those with

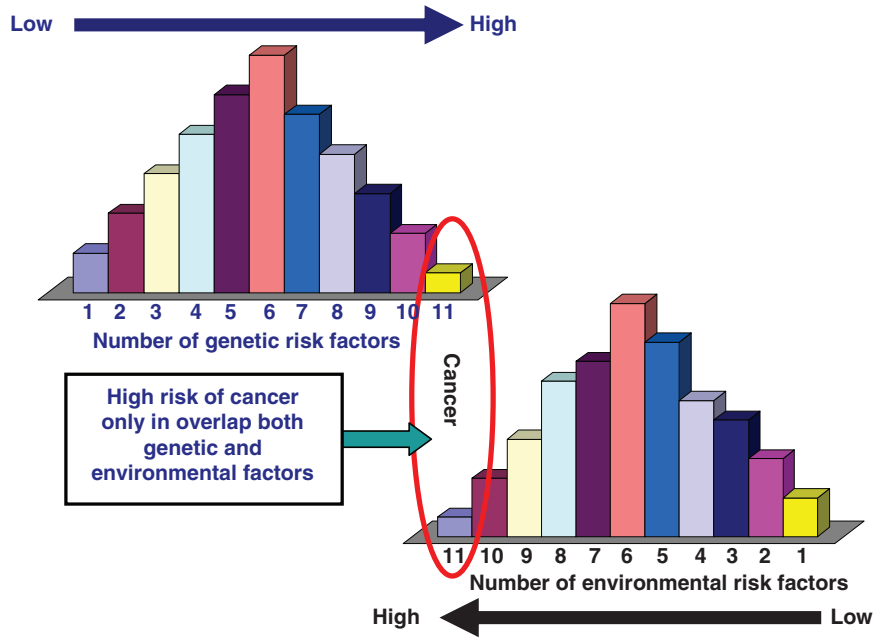


Fig. 8.1. Model of gene-environment interaction (modified from Talmud, 2007).

both the genotype and exposure of interest. The magnitude of an additive interaction effect was determined by estimating the interaction contrast ratio (ICR) and 95% confidence interval with the following formula: ICR = $OR_{cg} - OR_c - OR_g + 1$, where OR_{cg} is the OR for exposure with variant genotype, OR_c is the OR for exposure with a wide type genotype, and OR_g is the OR for variant genotype among nonexposed.

In fact, at the epidemiologic practice, epidemiologists take into account gene–environment interactions, in most situations, under dealing with two independent risk factors for cancer (the simplest model). The presence of two risk factors may increase cancer risk in a multiplicative or in an additive way. A theoretical example of gene–environment interaction is present in **Table 8.1**. Thus, interpretation of interaction depends on the underlying risk

Table 8.1
Theoretical example of gene–environment interaction[a]

	Odds ratios of genotype	
	Wild type	**Variant**
Unexposed	1.0(ref)	1.0
Exposed	3.0	5.0

[a]Interaction effect: multiplicative: 1.7 = 5/(3 × 1); additive: 2 = 5 – 3 –1 + 1.

model being used in the analysis. When the joint effect of genetic and environmental risk factors act through the same pathogenic pathway, the following statistical and biological perspectives have been introduced: 1) statistically, the interaction of two risk factors is simply the coefficient of the product term of the risk factors; and 2) biologically, interaction between two factors is defined as their coparticipation in the same causal mechanism leading to cancer development.

From a mechanistic viewpoint, the significant interaction suggests that the effect or by-product of the environmental insult modifies the molecular function of the product of the gene under observation at the molecular level. Statistical significance depends on the magnitude of the risk and the sample size. In large studies, small ORs such as 1.1 might be statistically significant but biologically meaningless. Such results could be due to unobserved variables (confounders). More complicated interaction models involved in multiple genetic and environmental factors are far from well developed.

At least six types of gene–environment interaction have been proposed theoretically (**Table 8.2**) assuming that the joint effect (interaction) of the environmental exposures and genetic factor always increase cancer risk ($OR_{cg} > 1$). In type 1, neither the environmental exposures nor the genetic risk factors have any effect individually, but interaction between them causes the cancer. In type 2 (considered the most important type of gene–environment interaction), the genetic risk factor has no effect on disease in the absence of exposure, but it exacerbates the effects of the exposure. Genetic susceptibility may be more important at low levels of exposure because at these dose levels, the relevant enzymes are saturated in both rapid and slow metabolizing, whereas this effect does not happen at high doses. Type 3 is the converse of the second that environmental exposure is ineffective per se, but it enhances the effect of genetic risk factor. Type 4 occurs when both environmental exposure and genetic risk factor increase the risk of cancer, and the combination is synergistic (i.e., skin cancer development due to effect of UV light on highly susceptible xeroderma pigmentosum patients defective in DNA nucleotide excision repair). In types 5 and 6, genetic factor is protective. The genetic factor is assumed to be protective in the absence of environmental exposure, but it becomes a risk in the presence of an environmental factor. The environmental factor alone has no effect on cancer risk in type 5, but has does have an influence in type 6.

Several examples may provide evidence that exploring gene–environment interactions are one of important pathways to better understand the role of risk factors for cancers *(4–6)*. Primary candidates for gene–environment interaction studies have been genes encoding enzymes involved in the metabolism of established cancer risk factors. Variant forms of these genes exist in populations,

Table 8.2
Six models of gene–environment interaction on cancers

Model 1	Neither the genetic factor nor the environmental exposure alone has any effect for cancer risk, but the interaction between them causes the cancer
Model 2	The genetic factor has no effect on cancer risk in the absence of environmental exposure, but it exacerbates the effect of an existing environmental risk factor
Model 3	The environmental exposure has no effect on cancer risk in the absence of genetic factor, but it enhances the effect of an already known genetic risk factor
Model 4	Both genetic and environmental factors influence cancer risk, and the combination is synergistic
Model 5	The genetic factor is assumed to be protective in the absence of environmental exposure, but become risk in the present of environmental factor. There is no effect of environmental exposure alone
Model 6	It is similar to model 5 except that environmental factor alone has influence on cancer risk

many of which alter metabolism and may thus increase or decrease exposure of humans to ultimate carcinogens. Because many of the variant gene forms are relatively common, interactions between carcinogen-metabolizing enzymes and common environmental factors may have a marked impact on the population attributable risk for cancer. Of particular interest in understanding the role of genes in cancer etiology is that these polymorphisms are often common and that their prevalence differs between populations *(7)*. Ethnic variation in cancer incidence and mortality may be due in part because of differences in the distribution of polymorphisms, as well as differences in environmental exposures and personal lifestyle and behavior.

In most cases, polymorphisms in enzymes involved in carcinogen metabolism seem to have little overall effect on cancer risk. A more interesting picture evolves from studies that have explored the effect of the polymorphisms within levels of environmental exposures. For example, *GSTM1* null is minimally

associated with lung cancer risk overall. Yet, the effect of *GSTM1* null on lung cancer seems increased among heavy smokers or nonsmokers exposed to high levels of environmental tobacco smoke. Indeed, in the presence of an interaction between genetic and environmental factors, failure to take into account the level of environmental exposure may lead to association studies producing conflicting results or lead to a biased estimation of disease risk.

Other studies, including ours, indicated that the effects of some genetic susceptibility profiles (in carcinogen metabolism and DNA repair pathways) are more easily observed in subjects with lower levels of exposure. We have observed a weak additive interaction between the *XRCC1* 399Gln allele (a polymorphism in base excision repair pathway) and detectable PAH-DNA adducts on breast cancer risk among never active smoking women, but no similar effects were noted in relation to active cigarette smokers. This gene–adduct interaction among never active smokers is consistent with observations among never or light smokers in other cancer sites (lung, bladder, and nonmelanoma skin cancer). A potential explanation is that cigarette smoking may stimulate DNA repair in response to the damage caused by tobacco carcinogens. Smoking or heavy tobacco exposure may lead to more proficient repair in lymphocytes compared with little or no smoke exposure. This might overwhelm any effect of the *XRCC1* codon 399 variant in smokers. Thus, the interaction observed in never active smokers may not mean that there is no interaction effect between PAH-DNA adducts and *XRCC1* codon 399 variant in smokers, only that it may be hard to detect. Thus, although there are many reports of gene–environment interactions on cancer risk, these results have to be viewed with caution and need validation because of the limited statistical power.

4. Study Designs in Evaluating Gene–Environment Interactions

Although cohort study and family study design can be used to evaluate the role of gene–environment interactions, practically, molecular epidemiologic studies currently mainly use two study designs to evaluate gene–environment interactions. One design is case-control study, which has been the hallmark design for traditional cancer epidemiology studies, and another is case-only study, a relatively new design. These designs have their unique strengths and limitations, but both require a sizeable number of study subjects to reliably evaluate gene–environment interactions.

4.1. Case-Control Study

Case-control study with population-based or hospital-based subjects is the most commonly used and efficient design in studies of gene–environment interaction. Similar to a traditional case-control study, the researcher samples subjects with cancer (cases), and an appropriate sample of subjects who are without cancer (controls). To avoid bias, it is important that the controls are selected from the same source population as the cases, especially because the prevalence of genetic polymorphisms varies greatly across populations. The prevalence of the polymorphism among cancer cases is compared with the prevalence among the controls, to measure the overall relative risk linking the polymorphism to cancer risk. To assess gene–environment interactions, the data can be stratified, so that one evaluates the association between the environmental exposure and cancer within different genotypes. By use of unexposed subjects with no susceptibility genotype as the referent group, ORs for all other groups can be estimated under either multiplicative or additive models. Adjustment for potential confounding variables may be accomplished by using stratification or multivariate approaches. Although both multiplicative and additive approaches can be used to assess gene–environment interactions in case-control studies, some researchers suggest that the multistage theory of cancer is better assessed by assuming departure from the additive model. The power of case-control study design for assessing interaction is depended on the frequencies of the environmental exposure and the genetic polymorphisms, which determine the number of cases and controls required. In fact, results from multiple studies suggest that case-control designs may be used to detect gene–environment interaction only when the environmental factor and genetic polymorphisms are common.

The main advantage of this design is that the main effects of the environmental exposure and genetic susceptibility as well as their interactive effect may be estimated. The main disadvantages are that this design may not be appropriate for the study of gene–environment interaction involving rare genes or uncommon environmental exposures. In addition, population stratification or genetic admixture might adversely influence this design.

4.2. Case-Only Study

Case-only study has become increasingly popular as a more efficient tool to study gene–environment interactions. In this design, the association between an environmental exposure and a genotype is examined among case subjects only. Cases are selected using epidemiologic principles of case selection as for any case-control study, with the exposure effect assessed among only cases. Cases carry wild-type genotype as the control group, and cases unexposed to the environmental factors serve as the referent group. Under the assumption of independence of exposure and

genotype, and the disease is rare, the OR of case-only is the interaction effect as measured in a traditional case-control study under a multiplicative model.

In the analysis, cases with a high-risk genotype are compared with cases without a high-risk genotype with respect to the prevalence of environmental exposures, and a relative risk (RR) is calculated. The following formula relates the RR from the case-only design: $RR_{CA} = RR_{GE}/(RR_G \times RR_E)$, where RR_{CA} is the case-only RR. For example, if the RR from the case-only study is 2.0, this indicates that the gene–environment interaction is two-fold higher than the product of the individual effects. Case-only studies offer better precision for estimating gene–environment interactions than those based on both cases and controls (i.e., there are smaller standard errors due to elimination of control group variability). The power for detecting gene–environment interactions in case-only studies is comparable with the power for assessing a main effect in a classical case-control study, i.e., the number of cases required is substantially less than the number that would be required in a case-control study to detect an interaction with the same power.

The main disadvantages of the case-only design are that the main effects of the genetic and environmental factors cannot be estimated and that the genetic and environmental factors must occur independently. However, the latter assumption cannot be assessed in this design. In addition, case-only studies cannot detect gene–environment interaction models with departures from an additive scale.

5. Bioinformatics and Detecting Gene–Environment Interactions

Because a complex disease such as cancer typically occurs through a multifactor interplay between many genetic and environmental factors, the effect of each individual factor is unlikely to be substantial. As the number of variables (genetic and environmental) increases when exploring multiple interactions, the number of interaction terms grows exponentially. This creates too many independent variables in relation to the number of outcome events (cases or controls). Because the number of possible interaction terms grows exponentially as each additional main effect is included in the model, logistic regression, like most parametric statistical methods, is limited in its ability to deal with interaction data involving many simultaneous factors. This may be a prohibitive factor when considering interactions between many candidate genes and environmental factors for complex diseases such as cancer. To

deal with these issues, many researchers are exploring variations and modifications of logistic regression such as logic regression, penalized logistic regression, and automated detection of informative combined effects. Additional explorations are being conducted in data reduction approaches and pattern recognition methods, including cluster analysis, cellular automata, support vector machines, self-organizing maps, systems biology, and neural networks *(7)*.

Multifactor dimensionality reduction (MDR), a data reduction approach for identifying gene–gene and gene–environment interaction effects, seeks to identify combinations of multilocus genotypes and discrete environmental factors that are associated with a high risk of cancer or a low risk of cancer. Thus, MDR is used to define a single variable that incorporates information from several loci, environmental factors, or both that can be divided into high-risk and low-risk combinations. This new variable can be evaluated for its ability to facilitate classification and prediction of cancer risk status by using cross-validation and permutation testing. Among all of the two-factor combinations, a single model that has the fewest misclassified individuals is selected. Thus, this two-locus model will have the minimum classification error among the two-locus models. To evaluate the predictive ability of the model, the prediction error is estimated using 10-fold cross-validation. The model with the combination of genetic factors, discrete environmental factors, or both that maximizes the cross-validation consistency and minimizes the prediction error is selected.

MDR method has demonstrated gene–gene interactions in a variety of different real data sets, including sporadic breast cancer, prostate cancer, hypertension, type II diabetes, and coronary artery calcification *(8)*. However, no gene–environment interaction has been confirmed so far by MDR. Although several new methods have been brought to the field, further explorations are needed to improve the detection of interactions.

Classification and regression tree (CART) analysis uses a binary recursive-partitioning method that identifies subgroups of high-risk subjects and detects higher order gene–gene and gene–environment interactions among a large number of variables. The advantages of CART include its ability to extract the most important variables from a large number of potential predictors and its ability to handle missing value in the data set. However, the CART analysis is a postdata mining tool, and the results should be interpreted with some degree of caution. External validation in an independent epidemiology study is warranted to confirm the potential high-risk groups identified by CART.

6. Implications in Public Health Practice

In theory, once the risk of particular combination of genotype and environmental exposure is known, medical interventions (including lifestyle advice, screening, or medication) could then be targeted at high-risk groups or individuals, with the aim of preventing cancer. The final goal of cancer epidemiologic study is also to prevent and decrease the occurrence of cancers at population level. Understanding the role of potential gene–environment interaction for cancer has special significance on cancer prevention. Each investigation that increases our understanding of gene–environment interaction, etiologic heterogeneity, pathogenesis, and natural history of common diseases adds to a knowledge base for estimating risks and guiding interventions to improve population health.

From the view of public health and cancer prevention, an individual is generally most concerned about voluntary environmental risks that can actually be controlled, such as tobacco, alcohol use, and obesity, rather than involuntary risks that cannot be controlled, such as radiation from a nuclear power plant or pesticides on food. Scientists also can easily communicate with every lay person that obesity or cigarette smoking is bad, eating fruit and vegetables is good, and drinking coffee is bad, or good, or does not matter to prevent cancers. Although the genetic factors, such as genetic polymorphisms, belong to the unchangeable factors, at least at current time, if some voluntary environmental factors have been confirmed to interact with certain genetic factors, people carrying these high-risk genetic factors may be more easily to be won over to change their lifestyle. Scientists can have more confidence to advise these people to lose weight, promote smoking cessation, eat cucumbers, and drink black coffee in moderation to decrease cancer risks *(9)*. Conversely, if we know only the etiologic role of genetic factors, such as most of the oncogenes and genetic variants unrelated with environmental factors by molecular research, we can do nothing. Neither removed these genetic factors nor modified them, to date at least. We cannot tell people to change a mutated gene or oncogenes that they already carry in their genome. Fortunately, with the development of biomedical science, several genetic or epigenetic biomarkers have been observed to be changeable according to different environmental status. It suggested that these genetic or epigenetic factors also may be preventable by controlling the related environmental factors, such as DNA hypermethylation may be affected by dietary polyphenols (epigallocatechin 3-gallate from green tea and genistein from soybean), and the rate of telomeres shorten

by chronic inflammation and oxidative stress. Thus, the study of gene–environment interaction not only provides a useful means to improve our understanding of cancer pathology at the molecular level but also allows specific targeting of prevention and therapies to high-risk subjects (those with a high-risk genotype or epigenetic status in a high-risk environment) *(10)*.

7. Future Perspectives

The common disease-common variant hypothesis have been turn out to be useful in exploring etiology of cancers. At the exciting time of epidemiologic research, efforts to improve both the efficiency and validity of evaluating gene–environment interactions will be of critical importance. These interactions can be additive, synergistic, or antagonistic. The integration of molecular biologic techniques into epidemiologic studies offers a plenty of opportunities to better understand these interactions.

One of perspectives of molecular epidemiology studies on cancer is to evaluate multiple biomarkers involved in both genetic and environmental factors in target tissues (biopsy or surgical) and surrogate biospecimens (peripheral blood) parallel at the same cancer cases or subjects with precancerous lesions. It will greatly increase the efficiency of the study design, and exclude potential false-positive results. Another perspective is the need for more epidemiologic evidence of exploring the interactions based on personal heterogeneity: exposures, susceptibility, and tumor types. The advent of whole genomic assessment of DNA, RNA, and proteins will enable the identification of biomarkers with heterogeneity. However, very large studies will be needed to permit stratification by these biomarkers of heterogeneity. Finally, further assessment of existing study designs and development of new approaches are needed to improve the ability to detect gene–environment interactions on cancer risk. The cohort studies to follow-up subjects with biomarkers in multiple points of life will greatly enhance our ability to evaluate the gene–environment interactions. As the knowledge for carcinogenesis has expanded rapidly, additional pathways with potential to interact with environmental factors will become the focus of molecular epidemiology research in the future, such as DNA repair, cell cycle control, apoptosis, genomic instability, immune response, and the inflammatory response etc. Searching for the potential interactions between genes in these pathways and environment factors will make molecular epidemiological studies more meaningful in cancer prevention and control.

References

1. Doll R, Peto R. (1981) The causes of cancer. J *Natl Cancer Inst* **66**, 1191–1308.
2. International Agency for Research on Cancer. (1990) Cancer: causes, occurrence and control. IARC, Lyon, France.
3. Perera FP, Weinstein IB. (2000) Molecular epidemiology: recent advances and future directions. *Carcinogenesis* **21**, 517–24.
4. Shields PG, Harris CC. (2000) Cancer risk and low penetrance susceptibility genes in gene-environment interactions. *J Clin Oncol* **18**, 2309–2315.
5. Khoury MJ, Wagener DK. (1993) Population and familial relative risks of disease associated with environmental factors in the presence of gene-environment interaction. *Am J Epidemiol* **137**, 1241–1250.
6. Goldstein AM, Andrieu N. (1999) Detection of interaction involving identified genes: available study designs. *J Natl Cancer Inst Monogr* **26**, 49–54.
7. Yang Q, Khoury M. (1997) Evolving methods in genetic epidemiology. III. Gene-environment interaction in epidemiologic research. *Epidemiol* Rev **19**, 33–43
8. Talmud PJ. (2007) Gene-environment interaction and its impact on coronary heart disease risk. *Nutr Metab Cardiovasc Dis* **17**, 148–52.
9. Khoury MJ, Davis RL, Gwinn M, Lindegren ML, Yoon PW. (2005) Do we need genomic research for the prevention of common diseases with environmental causes? *Am J Epidemiol* **161**, 799–805.
10. Ritchie MD. (2005) Bioinformatics approaches for detecting gene-gene and gene-environment interactions in studies of human disease. *Neurosurg Focus* **19**, E2.

Chapter 9

Introduction to the Use of Regression Models in Epidemiology

Ralf Bender

Summary

Regression modeling is one of the most important statistical techniques used in analytical epidemiology. By means of regression models the effect of one or several explanatory variables (e.g., exposures, subject characteristics, risk factors) on a response variable such as mortality or cancer can be investigated. From multiple regression models, adjusted effect estimates can be obtained that take the effect of potential confounders into account. Regression methods can be applied in all epidemiologic study designs so that they represent a universal tool for data analysis in epidemiology. Different kinds of regression models have been developed in dependence on the measurement scale of the response variable and the study design. The most important methods are linear regression for continuous outcomes, logistic regression for binary outcomes, Cox regression for time-to-event data, and Poisson regression for frequencies and rates. This chapter provides a nontechnical introduction to these regression models with illustrating examples from cancer research.

Key words: Regression, linear regression, logistic regression, Poisson regression, Cox regression.

1. Introduction

In analytical epidemiology, a common research question is whether an exposure such as radiation or dust has an impact on a response such as mortality or cancer. Frequently, the exposure is measured by means of a binary indicator (exposure yes/no). However, the use of more than two exposure categories as well as the use of a continuous variable to describe the exposure is also possible. In theory, the best study design to investigate the impact of an exposure on a response is given by a randomized controlled trial. However, in epidemiologic research randomization of people to different exposure groups is impossible in almost all cases. Thus,

Mukesh Verma (ed.), Methods in Molecular Biology, Cancer Epidemiology, vol. 471
© 2009 Humana Press, a part of Springer Science + Business Media, Totowa, NJ
Book doi: 10.1007/978-1-59745-416-2

common study designs in epidemiology to investigate relations between exposure and response are given by cohort studies and case-control studies. In these designs without randomization, the relation between two variables can be confounded by other variables. If this is not taken into account in data analysis, the results may be seriously biased.

Principally, there are two ways to deal with confounders in epidemiologic research: via study design or via data analysis. One example of a design-related method to take confounders into account is matching. One of the most important analysis related methods to deal with confounding is multiple regression analysis. Other methods are, for example, stratification, the use of instrumental variables, or the application of propensity scores. In this chapter, an overview of the most important multiple regression models is given with a focus on applications in modern epidemiology.

The term regression is somewhat misleading, and it does not describe the main feature of the method in its current applications. Historically, the term regression goes back to Galton who described the tendency for the offspring of seeds to be closer to the average than their parent seeds *(1)*. However, modern applications of regression methods do not only analyze such "regression to the mean" effects. Nowadays, regression methods are used to describe in general any relationship between a response variable, also called dependent or outcome variable and one or multiple explanatory variables, also called independent or predictor variables, covariates, or risk factors. In the case of only one explanatory variable, the corresponding model is called simple regression, the use of multiple explanatory variables leads to the application of multiple or multifactorial regression. Frequently, multiple regression is also called multivariate or multivariable regression. However, the use of these terms may be confusing because regression models dealing with multiple response variables are also called multivariate regression models *(2–5)*. To avoid confusion, we use the term multiple regression for models with one response and multiple explanatory variables. In this chapter, we discuss only regression models with one response variable. An introduction to multivariate regression models can be found in textbooks about multivariate statistical methods *(3,4)*.

Different regression models have been developed in dependence on the measurement scale of the response variable. The basic standard regression models are linear regression for continuous outcomes, logistic regression for binary outcomes, Cox regression for time-to-event data, and Poisson regression for frequencies and rates. In the next sections, the basic features of these types of regression models are summarized, followed by some remarks about model building. Finally, important extensions of the standard regression models are summarized.

2. Linear Regression

2.1. Simple Linear Regression

The use of simple linear regression is possible if the effect of one continuous explanatory variable X, for example, body mass index measured in kilograms per square meter on one continuous response variable Y, for example, systolic blood pressure measured in millimeters of Hg is to be investigated. The fundamental model equation is given by

$$Y = \beta_0 + \beta_1 x + e \tag{1}$$

where β_0 is called intercept, β_1 is the regression coefficient for X, x is the observed value of X, and e is the random error describing individual deviations from the mean of Y given X = x, also called residual term. Through the functional form a straight line is described, where the intercept β_0 represents the mean of Y for x = 0 and the regression coefficient β_1 represents the slope of the line, that is, the average increase of Y for a one-unit increase of X. The fundamental assumption of simple linear regression is that the true association of Y and X is in fact linear. This assumption has to be fulfilled at least approximately to yield interpretable results. If the association of Y and X cannot be approximated by a straight line, more complicated models should be applied (see subheadings 7.4. and 7.5.). Another usual assumption is that the residual term is normally distributed with mean 0 and variance σ^2. However, the assumption of normally distributed residuals is negligible in practice if the sample size is large enough and the distribution of the residual term is not too skew.

In epidemiologic practice, the parameters β_0 and β_1 have to be estimated from data of n subjects that should represent a random sample of the considered population. In the simplest case, the data consist of n independent pairs (x_i,y_i), i=1,...,n, obtained from n subjects, where x_i and y_i represent the observed values of subject i for X and Y, respectively. In the case of repeated measurements on the same subject or other reasons leading to correlated data, the use of more complicated models is required (see subheading 7.6.). In the case of n independent pairs (x_i,y_i) for i=1,...,n, the parameters β_0 and β_1 of model 1 can be estimated from these data. The standard estimation method is given by the method of ordinary least squares (OLS), which results in a fitted regression line that minimizes the average of the squared deviation of the line from the observed data *(1)*. The OLS estimates of the parameters β_1 and β_0 are given by

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{2}$$

where $\bar{x}$ and $\bar{y}$ are the arithmetic means of x_i and y_i, i=1,…,n, respectively.

Before a straight line is fitted to data they should be examined graphically by means of a scatter plot of Y versus X to assess whether the linearity assumption holds at least approximately. If this is the case, the estimated effect of X on Y can be described by the estimated slope, because the larger β_1 the steeper the regression line. The estimation uncertainty should be described by means of the standard error (SE) and the 95% confidence interval (CI) for β_1. It is also possible to test the null hypothesis H_0: $\beta_1=0$ by means of a *t* test. Only in the case of a significant test result we can conclude that a significant effect of X on Y is found. Additionally, we can estimate the expected value of Y for given values of X. Again, the estimation uncertainty should be described by means of SEs and CIs. For the whole fitted line confidence bands are useful.

In summary, the goals of regression models are threefold:

1. To determine whether the variables Y and X are systematically related (test of H_0: $\beta_1=0$).
2. To estimate the effect size of X on Y by means of $\hat{\beta}_1$ (complemented by a 95% CI).
3. To predict the expected value of Y for given values of X (with 95% CI).

2.2. Multiple Linear Regression

In epidemiology, simple linear regression plays only a negligible role. First, binary and time to event outcomes are much more common than continuous response variables (see subheadings 3.4., and 5.). Second, in almost all cases, the effects of confounders have to be taken into account, so that multiple linear regression should be applied even if we are interested mainly in the effect of one primary explanatory variable. The multiple linear regression model is an extension of model 1 where instead of only one explanatory variable X several explanatory variables $X_1,\ldots,X_k$ with observed values $x_1,\ldots,x_k$ are considered. The fundamental model equation is given by

$$Y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + e \qquad (3)$$

where β_j represent the regression coefficients for X_j for j=1,…,k. It is not required that all of the explanatory variables $X_1,\ldots,X_k$ are measured on a continuous scale. It is also possible to include binary explanatory variables (e.g., exposure yes/no). In this case, we should choose one reference group (e.g., unexposed subjects), assign the value 0 to this group, and assign the value 1 to the other group (e.g., exposed subjects). Then, the corresponding regression coefficient describes the mean response difference of exposed and unexposed subjects. It is also possible to include categorical explanatory variables with more than two groups. One

possibility is to create so-called dummy variables. For example, if the categorical variable X is given by marital status with the categories single, married, divorced, and widowed, we should choose one reference group (e.g., single) and create the new binary dummy variables X_1=married (yes/no), X_2=divorced (yes/no), and X_3=widowed (yes/no). In general, a categorical explanatory variable with m categories can be included in a multiple regression model by means of m–1 dummy variables. Thus, model 3 represents a general tool to assess the effects of continuous and categorical explanatory variables on a continuous response variable. As a consequence, other well-known statistical methods like unpaired *t* test and analysis of variance are embedded in the general multiple linear regression model 3. For example, the test of H_0: $\beta_1=0$ is equivalent to the unpaired *t* test, if the model contains only one binary explanatory variable.

As discussed above, the usual method for parameter estimation is given by OLS. The corresponding formulas are best summarized by using matrix notation, which goes beyond the scope of this chapter. The interested reader can find the formulas for example in Matthews *(6)* or standard textbooks *(7,8)*. As discussed above, we can test whether the explanatory variable X_j (j=1,…,k) has a significant effect on the response Y, we can describe the estimated effect size of X_j by means of $\hat{\beta}_j$ for j=1,…,k (with CIs), and we can predict the expected value of Y for given values of $X_1,\ldots,X_k$ (with CIs). The advantages of model 3 in comparison with model 1 are given by the fact that model 3 can describe the effects of several explanatory variables simultaneously and that the regression coefficient β_j represents the effect for X_j adjusted for all other explanatory variables. In general, it is misleading to assess the associations of the response Y and several explanatory variables by means of several simple linear regression models. The corresponding simple estimates may be seriously biased, because the effects of other explanatory variables are not taken into account.

2.3. Interactions

In addition to the assumptions of model 1, model 3 contains another important assumption. It is assumed that the effects of the explanatory variables $X_1,\ldots,X_k$ are additive, i.e., the effect of one explanatory variable is independent of all other explanatory variables. If this is not the case, the model should be extended by inclusion of appropriate interaction terms. This can be best explained by using two binary explanatory variables, say X_1=sex (male=1, female=0) and X_2=exposure (yes=1, no=0). If the effect of exposure is different for men and women, the model should contain the corresponding interaction term given by $X_3=X_1\times X_2$, that means X_3=1 for exposed men and X_3=0 otherwise. If we include the interaction term X_3 in the regression model, the corresponding regression coefficient describes the additional

effect of the exposure in males. In detail, β_0 represents the mean response for unexposed females, β_1 the mean response difference between unexposed males and females (sex effect in unexposed subjects), β_2 the mean response difference between exposed and unexposed females (exposure effect in females), $\beta_2+\beta_3$ the mean response difference between exposed and unexposed males (exposure effect in males), and $\beta_1+\beta_3$ the mean response difference between exposed males and females (sex effect in exposed subjects). In this model, there are no unique effects, neither for sex nor for exposure. The effect of sex is dependent on the exposure status, and the effect of exposure is dependent on sex.

2.4. Coefficient of Determination

A frequently used measure to describe the predictive ability of a linear regression model is the coefficient of determination, R^2, defined by

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad (4)$$

where

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \ldots + \hat{\beta}_1 x_{ki} \quad (5)$$

is the estimated response value for subject i with values $x_{1i},\ldots,x_{ki}$ for the explanatory variables. R^2 is the proportion of the model sums of squares to the total sums of squares and represents the proportion of variability of the response explained by the explanatory variables all together. Frequently, in practice R^2 is low although the model contains one or more significant explanatory variables with very low *P* values, especially if the sample size is large. A number of other measures and graphical tools to describe the goodness-of-fit of linear regression models are available *(7,9)*.

2.5. Remarks

Multiple linear regression can be applied in cohort studies and in cross-sectional studies. In all cases, one should be careful in interpreting the results, because significant regression coefficients demonstrate statistical associations but not necessarily causal effects. An additional major limitation of cross-sectional studies is that, in general, it cannot be clearly determined whether an explanatory variable has an effect on the response or vice versa.

2.6. Example

Chan et al. *(10)* applied multiple linear regression to estimate standard liver weight for assessing adequacies of graft size in live donor liver transplantation and remnant liver in major hepatectomy for cancer. Standard liver weight (SLW) in grams, body weight (BW) in kilograms, gender (male=1, female=0), and other anthropometric data of 159 Chinese liver donors who underwent donor right hepatectomy were analyzed. The formula

$$SLW = 218 + 12.3 \times BW + 51 \times gender \quad (6)$$

was developed and a coefficient of determination of R^2=0.48 was reported *(10)*. These results mean that in Chinese people, on average, for each 1-kg increase of BW, SLW increases about 12.3 g, and, on average, men have a 51-g higher SLW than women. Unfortunately, SEs and CIs for the estimated regression coefficients were not reported. By means of formula 6 the SLW for Chinese liver donors can be estimated if BW and gender are known. About 50% of the variance of SLW is explained by BW and gender.

3. Logistic Regression

3.1. Importance of Logistic Regression

In cancer epidemiology, frequently the associations of risk factors with development of cancer or cancer mortality are investigated. This leads to binary response variables (e.g., cancer yes/no). The usual regression model for binary responses is logistic regression. It has been shown that this model can not only be applied in cohort studies but also in case-control studies *(11,12)*. This makes logistic regression one of the most important statistical methods in epidemiologic research. Levy and Stolte found that in about one of three articles published in the *American Journal of Public Health* and the *American Journal of Epidemiology* between 1990 and 1998 logistic regression was used *(13)*.

3.2. Explanation of Logistic Regression

The basic question leading to logistic regression is the same as in subheading 2. about linear regression. Are there associations between the explanatory variables $X_1,\ldots,X_k$ and the response variable Y? The only difference to the data situation described in subheading 2. is that the response Y is a binary variable (event yes/no). The naive use of linear regression would mean that we fit a continuous relation between the explanatory variables and the response, although there are only two possible values for Y, namely, 0 and 1. However, if we look at the event probability π=P(Y=1), we have a continuous term with possible values between 0 and 1, which can be modeled in a continuous manner, but it has the inconvenient restriction that values lower than 0 and higher than 1 are impossible. If we look at the odds $\pi/(1-\pi)$, we have a continuous term with possible values between 0 and infinity and if we take the natural logarithm of the odds $\log(\pi/(1-\pi))$, called logit, we have a mathematically convenient term with no restrictions of the domain. The logistic regression model is given by the linear relation between the logit and the values of the explanatory variables,

$$\log(\pi/(1-\pi)) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k \quad (7)$$

Equation 7 is mathematically equivalent to

$$\pi = \frac{\exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k)} \quad (8)$$

The mathematical function described in formula 8 is called logistic function explaining the term logistic regression. Fortunately, the basic features of multiple regression as described in subheading 2. (need to check model assumptions, use of dummy variables, use of interaction terms) also apply to logistic regression. Thus, we can concentrate on the differences between logistic and linear regression.

Because the logits are not available for each individual, we cannot simply create scatter plots between the response and the explanatory variables as in linear regression. We have to use percentage plots based on grouped data as described by Hosmer and Lemeshow *(14)* to assess the relation between the event probability and the explanatory variables graphically.

For the same reason, it is not possible to use OLS for parameter estimation in logistic regression. However, parameter estimation can be performed by means of maximum likelihood, which is available in all statistical program packages containing logistic regression.

Several special tools are available to assess the goodness-of-fit and predictive ability of multiple logistic regression models *(14)*. One frequently used method is the Hosmer–Lemeshow test *(15)*.

3.3. Adjusted Odds Ratios (ORs)

The effect size of X_j on the response in logistic regression is best described by means of $\exp(\beta_j)$, because in model 7 it can be shown that

$$\exp(\beta_j) = OR_j \quad (9)$$

where OR_j represents the OR for X_j adjusted for the other explanatory variables. However, the simple equation 9 is only valid in models without interactions. If the model contains interaction terms the OR of one variable depends on the values of other explanatory variables, that is, it is impossible to describe the effect of a variable by means of only one unique OR. In the case of a continuous explanatory variable X_j (and a model without interactions), OR_j according to equation 9 describes the factor by which the odds of an event changes for each one-unit increase of X_j. Thus, the presentation of estimated ORs based upon logistic regression with continuous explanatory variables is only meaningful if the unit of the explanatory variables is known.

If the considered event is rare (risk lower than 10%), ORs can be interpreted as risk ratios because the corresponding numbers are approximately identical. This is usually valid in case-control studies. However, in cohort studies investigating common diseases, ORs cannot be interpreted as risk ratios, because the absolute value of ORs is larger than that of risk ratios. In this case the reporting of the results in terms of risk ratios requires additional calculations *(16)*.

3.4. Example

Gorini et al. *(17)* investigated the association between alcohol intake and Hodkin's lymphoma by means of logistic regression. A population-based case-control study of 363 Hodkin's lymphoma cases and 1,771 controls was conducted yielding information about sociodemographic characteristics, tobacco, and alcohol consumption. Because a significant interaction between smoking and alcohol consumption was found separate analyses for nonsmokers and ever-smokers were performed taking the potential confounders gender, age, area of residence, education level, and type of interview into account. We consider only the results for nonsmokers. For the alcohol intake categories 0.1–9.0, 9.1–17.9, 18.0–31.7, and >31.7 g/day, the adjusted ORs (95% CIs) 0.45 (0.27–0.74), 0.52 (0.30–0.90), 0.27 (0.13–0.57), and 0.35 (0.15–0.79) compared with never-drinkers were reported *(17)*. Overall, the adjusted OR for ever-drinkers compared with never-drinkers was OR=0.46. This result means that in nonsmokers alcohol consumption has a protective effect because the risk of Hodkin's lymphoma is approximately halved in ever-drinkers compared with never-drinkers. However, a clear decreasing trend of the risk with increasing alcohol consumption could not be found.

4. Cox Regression

4.1. Time to Event Data

Logistic regression can be used when time is not relevant for the considered event or the risk of the considered event refers to a fixed time interval, for example 5 years. In cohort studies and clinical trials, frequently the participants are observed for different time periods due to staggered entry or censoring. Those designs lead to time to event data, also called survival or failure time data, where the time T until an event occurs is used as response rather than a binary variable Y. The distribution of time to event data is mostly described by the survival distribution function $S(t)=P(T \geq t)$ estimated by the method of Kaplan and Meier *(18)*. The impact of a categorical explanatory variable can be assessed by estimating Kaplan–Meier curves for each group. However, to assess the association of several possibly continuous explanatory

variables $X_1,\ldots,X_k$ with the response variable T a multiple regression model is required. For analyzing time to event data, frequently the Cox regression model is the first considered approach, and this method is probably the most widely used statistical model in general medical research *(19)*.

4.2. Modeling of Hazard Function

The essential idea of Cox *(20)* was to model the hazard function rather than the mean of T in dependence on the values of the explanatory variables. The hazard function of a survival time variable T is defined as

$$\lambda(t) = \lim_{\delta t \to 0} \frac{P(T \le t + \delta t \mid T \ge t)}{\delta t} \qquad (10)$$

Other names for the hazard function are hazard rate, intensity function, instantaneous death rate, and force of mortality. Roughly speaking, the hazard function is the probability that a person who is alive at time t will die in the next moment after t per unit of time. Cox proposed to model the hazard function as product of an arbitrary unspecified baseline hazard $\lambda_0(t)$ and an exponential term that is linear in the values of the explanatory variables $X_1,\ldots,X_k$

$$\lambda(t) = \lambda_0(t)\exp(\beta_1 x_1 + \ldots + \beta_k x_k) \qquad (11)$$

The baseline hazard $\lambda_0(t)$ represents the hazard function for an individual with $x_1=\ldots=x_k=0$. Model 11 forces the hazard ratio (HR) of two persons to be constant over time. Therefore, model 11 is also called proportional hazards model.

Due to the unspecified baseline hazard $\lambda_0(t)$ the Cox model is a semiparametric model in which full maximum likelihood estimation is not possible. However, parameter estimation can be performed by means of the so-called maximum partial likelihood estimation *(21)*.

Several graphical tools and formal tests have been proposed to investigate the goodness-of-fit of Cox models, especially to assess the adequacy of the proportional hazards assumption. A comprehensive overview is presented by Sasieni *(19)*.

4.3. Adjusted HRs

An important feature of the Cox model (without interactions) is that

$$\exp(\beta_j) = HR_j \qquad (12)$$

where HR_j represents the hazard ratio for X_j adjusted for the other explanatory variables. For a binary explanatory variable HR is the ratio of the hazards between the two groups and for a continuous explanatory variable HR represents the factor by which the hazard changes for each one-unit increase of the explanatory variable. As discussed before, in models containing interactions it

is not possible to describe the effect of an explanatory variable by means of one unique effect measure.

4.4. Example

The Cox regression model was used in the Seven Countries Study to analyze the association between cigarette smoking and total and cause-specific mortality *(22)*. In short, the Seven Countries Study is a longitudinal observational study of risk factors for coronary heart disease (CHD) in 16 cohorts situated in seven countries (Europe, United States, and Japan). The baseline examination of 12,763 men aged 40 to 59 years took place between 1957 and 1964. In the analysis of the 25-year follow-up data for the outcomes total mortality, CHD mortality, and lung cancer mortality the adjusted hazard ratios per 20 cigarettes/day 1.7 (p<0.001), 1.7 (p<0.001), and 4.2 (p<0.001) were reported *(22)*. In the corresponding Cox models baseline cohort residence, body mass index, serum cholesterol, systolic blood pressure, and presence of clinical cardiovascular disease were included as covariates. These findings confirm that smoking is a risk factor for total, CHD, and lung cancer mortality.

5. Poisson Regression

5.1. Modeling of Counts

For a response variable Y that represents counts or frequencies it is frequently reasonable to assume that the logarithm of the mean μ of Y is linearly related to the values of the explanatory variables $X_1,\ldots,X_k$, i.e.,

$$\log(\mu) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k \qquad (13)$$

which is equivalent to an exponential relationship between the mean and the explanatory variables

$$\mu = \exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k) = \exp(\beta_0) \times \exp(\beta_1) x_1 \times \ldots \times \exp(\beta_k) X_k \qquad (14)$$

Thus, a one-unit increase of X_j has a multiplicative effect given by $\exp(\beta_j)$ on the expected frequency μ.

Model 13 is called Poisson regression because maximum likelihood estimation of the regression coefficients can be performed based upon the assumption that the response variable Y is Poisson distributed *(23)*. The so-called log-linear models for contingency tables represent a special case of model 13 where only categorical explanatory variables are considered.

5.2. Modeling of Rates

Poisson regression models play an important role in epidemiologic research, because they also can be used in cohort studies to

analyze time to event data. Here, one makes use of the so-called Poisson process, in which the waiting times between successive events are independent and exponentially distributed with mean $1/\lambda$. Then, the number $Y(t)$ of events that occur up to time t follows a Poisson distribution with mean $\mu=\lambda t$. Note that the mean of $Y(t)/t$ equals λ. Applying model 13 for λ

$$\log(\lambda) = \log(\mu / t) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k \qquad (15)$$

is equivalent to

$$\log(\mu) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \log(t) \qquad (16)$$

where $\log(t)$ is called offset, because it can be considered as explanatory variable with regression coefficient set equal to 1.

Model 15 also can be written as

$$\lambda = \exp(\beta_0) \times \exp(\beta_1 x_1 + \ldots + \beta_k x_k) \qquad (17)$$

By defining $\lambda_0=\exp(\beta_0)$ as baseline rate we should notice the similarity between formulas 17 and 11. For a reasonable application of Poisson regression to time to event data, it is required to organize the data in an event-time table defined by a cross-classification over a set of time intervals and categories of the explanatory variables. A relatively fine stratification should be used because it is assumed that the hazard in each cell is constant. This could mean that the table is based on individual subjects and the only grouping refers to small time units in which approximately an exponential distribution can be assumed *(24)*.

The Poisson regression approach represents an important alternative to the Cox model because it is an efficient and intuitive method for handling of cumulative exposures, for allowing dependence of risk on multiple time scales, for direct estimation of the survival distribution function, and for direct estimation of relative risks and risk differences *(24,25)*.

An important goodness-of-fit measure suitable for Poisson regression models is given by the deviance, which compares the likelihood for the model that perfectly fits the data with the likelihood of the model under consideration *(26)*. Other methods are summarized by Seeber *(26)*. However, the usefulness of common methods to assess goodness-of-fit of Poisson regression models is limited in the case of large detailed event-time tables *(24)*.

5.3. Example

Romundstad et al. *(27)* used Poisson regression to analyze cancer incidence in a cohort of 2,620 male workers in the Norwegian silicon carbide industry. Cumulative exposures to total dust, silicon carbide fibers, and others were assessed by means of a job-exposure matrix. In Poisson regression models age and calendar period of diagnosis were taken into account. Concerning the outcome lung cancer and the exposure groups 0.1–0.9, 1.0–4.9, and ≥5

silicon carbide fibers per ml/year the adjusted risk ratios (95% CI) 3.0 (1.6–5.9), 2.9 (1.4–6.0), and 4.4 (2.1–9.0) compared with unexposed workers were reported *(27)*. These findings suggested an increased risk of lung cancer in workers of the silicon carbide industry due to work related agents such as silicon carbide fibers.

6. Model Building

In subheadings 2. to 5. the basic features of regression models are summarized focusing on the most important application areas in epidemiology and the interpretation of the main results. To apply these models adequately in practice, fundamental statistical knowledge about model building is required. In short, model building consists of the following interactive steps: choice of dependent and explanatory variables, model fitting, and model checking. Important issues in model building are given by the functional form of the considered variables (*see* also **Subheading 7.**), interactions, correlation among explanatory variables, outliers, sample size, and number of events for binary and time to event data. In general, a useful model can be obtained if an appropriate balance between goodness-of-fit, simplicity, and consistency with subject-matter knowledge is possible *(28)*. The choice of a model building strategy depends on the aim of the analysis and the specific data situation. More detailed guidelines how to use regression methods in practice are given in other textbooks and articles *(2,7,8,14,29–32)*.

7. Model Extensions

Only the basic standard formulations of the most important types of regression models are described so far. Several generalizations and extensions have been developed allowing more flexible applications. In the following sections, a short overview with main references is given.

7.1. Generalized Linear Models

At first, it should be noted that linear regression, logistic regression, and Poisson regression are special types of the class of generalized linear models (GLMs) *(2)*. Let μ be the mean of the response variable Y and $x_1,\ldots,x_k$ the observed values of the explanatory variables $X_1,\ldots,X_k$. Then, the GLM is defined by

$$g(\mu) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k \qquad (18)$$

where $g(\cdot)$ is called link function. For continuous response variables the use of the identity link leads to linear regression. For categorical responses, logistic regression and Poisson regression are obtained by using the logit link and the log link, respectively. The use of other link functions leads to other models. For example, in dose-response studies the so-called probit link, which represents the inverse of the cumulative distribution function of the standard normal distribution, is frequently used. The corresponding model is called probit regression. A comprehensive overview of GLMs is given by McCullagh and Nelder *(2)*.

7.2. Extensions of Binary Logistic Regression

As mentioned, logistic regression can be applied in cohort studies as well as in case-control studies. For matched case-control studies, however, a special procedure is required, the so-called conditional logistic regression *(14,33)*. Further extensions of the standard binary logistic regression have been developed to analyze nominal and ordinal response variables. The most important models are given by polytomous logistic regression *(34)* and the proportional odds model *(35)*. A nontechnical introduction to regression models for ordinal data is given by Bender and Grouven *(36)*, more comprehensive overviews are provided by Bender and Benner *(37)* and in several textbooks *(2,8,14)*.

7.3. Extensions of Cox Model

The most important generalizations of the Cox model are given by stratified Cox models and Cox models with time-dependent covariates *(30)*. Stratified Cox models allow for separate baseline hazards for different groups and can be used to take categorical covariates into account for which the proportional hazards assumption is not fulfilled. Cox models including time-dependent covariates can be used in situations in which the values of explanatory variables change over time. However, caution is advised for model fitting and interpretation of the results *(38,39)*. Especially in radiation research, besides the Cox model 11, which represents a multiplicative model for the hazard function, also additive hazard models are frequently used *(24,40)*. It is also common to incorporate external standard rates in the analysis leading to standardized mortality ratio regression *(40)*. Recent research focuses on developing additive hazard models including external standard rates to estimate relative survival in population-based cancer studies *(41)*.

7.4. Fractional Polyomials

One basic assumption of the presented regression models is linearity, that is, the assumption that the outcome is linearly related to all explanatory variables $X_1,\ldots,X_k$. This strong assumption is frequently violated in practice so that a more flexible approach is required. One possibility is to use transformations of the

explanatory variables. A useful approach that is developed to deal simultaneously with several continuous explanatory variables is given by fractional polynomials *(42,43)*. Useful introductions to this approach for epidemiologic applications are available *(44,45)* as well as a description of software tools *(46)*.

7.5. Nonlinear Models

Although possibly nonlinear transformations of explanatory variables such as fractional polynomials are used, all regression models discussed so far belong to the class of linear models in the sense that they are linear in the regression coefficients after appropriate transformation of the response. Sometimes relations between explanatory variables and response variables cannot be adequately described by linear models, but models that are nonlinear in the regression coefficients are suitable. These methods are called nonlinear regression models, and they are frequently used for example in pharmacokinetics. An overview of nonlinear regression models is given in several textbooks *(47–49)*.

7.6. Clustered and Correlated Data

Another basic assumption of the standard regression models is that the individual observations used to estimate the regression coefficients are independent. However, many designs and data situations lead to correlated response data, for example, cross-over trials, cluster randomized trials, or repeated measurements over time. In case of correlated or clustered data, the application of standard regression models is in general invalid and leads to incorrect *P* values and CIs. In short, there are two common ways to deal adequately with correlated data, both represent extensions of the class of GLMs. In the generalized estimating equations (GEE) approach *(50)*, the usual formulation of regression models is used. However, the estimation procedure accounts for correlation and a robust estimation of SEs and CIs is used *(51,52)*. An alternative to GEE modeling is given by generalized linear mixed models (GLMMs) *(53)*, also called random effects models, random coefficient models, multilevel models, or hierarchical models. In the context of survival time data the term frailty model is used *(54)*. Mixed models extend standard regression models by adding random effects. A random effect is permitted to vary from subject to subject or from cluster to cluster leading to a specific correlation structure of the considered data. A comprehensive overview of GLMMs is given by Brown and Prescott *(53)*, a nontechnical introduction in the context of public health research is provided by Diez-Roux *(55)*.

7.7. Missing Data and Measurement Error

Finally, two frequent complications in practical applications of regression models in epidemiology should be mentioned: missing values and measurement error. In standard applications of regression analyses it is assumed that all data are available and all data are measured without error. Frequently, however, more or less data

values are missing or some explanatory variables can be observed only with substantial measurement error. Ignoring missing values and measurement error could lead to substantial bias. Thus, it is important to investigate the reasons leading to missing data or measurement error. An overview of techniques to deal with missing values are given by Little and Rubin *(56)*, regression methods accounting for measurement error are described by Carroll et al. *(57)*.

References

1. Matthews DE. (2005). Linear regression, simple. In: Armitage P, Colton T, eds. Encyclopedia of Biostatistics, 2nd ed., vol. 4. Chichester, UK: Wiley, pp. 2812–2816.
2. McCullagh P, Nelder JA. (1989). Generalized Linear Models, 2nd ed. New York: Chapman & Hall.
3. Srivastava MS. (2002). Methods of Multivariate Statistics. New York: Wiley.
4. Anderson TW. (2003). An Introduction to Multivariate Statistical Analysis, 3rd ed. New York: Wiley.
5. Krzanowski WJ. (2005). Multivariate multiple regression. In: Armitage P, Colton T, eds. Encyclopedia of Biostatistics, 2nd ed., vol. 5. Chichester, UK: Wiley, pp. 3552–3553.
6. Matthews DE. (2005). Multiple linear regression. In: Armitage P, Colton T, eds. Encyclopedia of Biostatistics, 2nd ed., vol. 5. Chichester, UK: Wiley, pp. 3428–3441.
7. Draper NR, Smith H. (1998). Applied Regression Analysis, 3rd ed. New York: Wiley.
8. Harrell FE Jr. (2001). Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. New York: Springer.
9. Cook DR, Weisberg S. (1997). Graphics for assessing the adequacy of regression models. J Am Stat Assoc **92,** 490–499.
10. Chan SC, Liu CL, Lo CM, et al. (2006). Estimating liver weight of adults by body weight and gender. World J Gastroenterol **12,** 2217–2222.
11. Anderson JA. (1972). Separate sample logistic discrimination. Biometrika **59,** 19–35.
12. Mantel N. (1973). Synthetic retrospective studies and related topics. Biometrics **29,** 479–486.
13. Levy PS, Stolte K. (2000). Statistical methods in public health and epidemiology: a look at the recent past and projections for the next decade. Stat Methods Med Res **9,** 41–55.
14. Hosmer DW Jr, Lemeshow S. (2000). Applied Logistic Regression, 2nd ed. New York: Wiley.
15. Hosmer DW, Lemeshow S. (1980). Goodness-of-fit tests for the multiple logistic regression model. Commun Stat Theory Methods **9,** 1043–1069.
16. Davies HTO, Crombie IK, Tavakoli M. (1998). When can odds ratios mislead? BMJ **316,** 989–991.
17. Gorini G, Stagnaro E, Fontana V, et al. (2007). Alcohol consumption and risk of Hodgkin's lymphoma and multiple myeloma: a multicentre case-control study. Ann Oncol **18,** 143–148.
18. Kaplan EL, Meier P. (1958). Nonparametric estimator from incomplete observations. J Am Stat Assoc **53,** 457–481.
19. Sasieni P. (2005). Cox regression model. In: Armitage P, Colton T, eds. Encyclopedia of Biostatistics, 2nd ed., vol. 2. Chichester, UK: Wiley, pp. 1280–1294.
20. Cox DR. (1972). Regression models and life tables (with discussion). J R Stat Soc B **34,** 187–220.
21. Cox DR. (1975). Partial likelihood. Biometrika **62,** 269–276.
22. Jacobs DR Jr, Adachi H, Mulder I, et al. (1999). Cigarette smoking and mortality risk: twenty-five-year follow-up of the Seven Countries Study. Arch Intern Med **159,** 733–740.
23. Frome EL, Kutner MH, Beauchamp JJ. (1973). Regression analysis of Poisson-distributed data. J Am Stat Assoc **68,** 935–940.
24. Preston DL. (2005). Poisson regression in epidemiology. In: Armitage P, Colton T, eds. Encyclopedia of Biostatistics, 2nd ed., vol. 6. Chichester, UK: Wiley, pp. 4124–4127.
25. Spiegelman D, Hertzmark E. (2005). Easy SAS calculations for risk or prevalence ratios and differences. Am J Epidemiol **162,** 199–200.
26. Seeber GUH. (2005). Poisson regression. In: Armitage P, Colton T, eds. Encyclopedia of Biostatistics, 2nd ed., vol. 6. Chichester, UK: Wiley, pp. 4115–4124.

27. Romundstad P, Andersen A, Haldorsen T. (2001). Cancer incidence among workers in the Norwegian silicon carbide industry. Am J Epidemiol **153,** 978–986.
28. Royston P. (2000). A strategy for modelling the effect of a continuous covariate in medicine and epidemiology. Stat Med **19,** 1831–1847.
29. Harrell FE Jr, Lee KL, Mark DB. (1996). Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med **15,** 361–387.
30. Hosmer DW Jr, Lemeshow S. (1999). Applied Survival Analysis: Regression Modelling of Time to Event Data. New York: Wiley.
31. Bagley SC, White H, Golomb BA. (2001). Logistic regression in the medical literature: standards for use and reporting, with particular attention to one medical domain. J Clin Epidemiol **54,** 979–985.
32. Katz MH. (2003). Multivariable analysis: A primer for readers of medical research. N Engl J Med **138,** 644–650.
33. Breslow NE, Day NE. (1980). Statistical Methods in Cancer Research Vol. I: The Analysis of Case-Control Studies. Lyon, France: International Agency for Research on Cancer.
34. Engel J. (1988). Polytomous logistic regression. Stat Neerl 42: 233–252.
35. McCullagh P. (1980). Regression models for ordinal data (with discussion). J R Stat Soc B **42,** 109–142.
36. Bender R, Grouven U. (1997). Ordinal logistic regression in medical research. J R Coll Physicians Lond **31,** 546–551.
37. Bender R, Benner A. (2000). Calculating ordinal regression models in SAS and S-Plus. Biom J **42,** 677–699.
38. Andersen PK. (1992). Repeated assessment of risk factors in survival analysis. Stat Methods Med Res **1,** 297–315.
39. Altman DG, De Stavola BL. (1994). Practical problems in fitting a proportional hazards model to data with updated measurements of the covariates. Stat Med **13,** 301–341.
40. Breslow NE, Day NE. (1987). Statistical Methods in Cancer Research Vol. II: The Design and Analysis of Cohort Studies. Lyon, France: International Agency for Research on Cancer.
41. Dickman PW, Sloggett A, Hills M, Hakulinen T. (2004). Regression models for relative survival. Stat Med **23,** 51–64.
42. Royston P, Altman DG. (1994). Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. Appl Stat **43,** 429–467.
43. Sauerbrei W, Royston P. (1999). Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials. J R Stat Society **162,** 71–94.
44. Royston P, Ambler G, Sauerbrei W. (1999). The use of fractional polynomials to model continuous risk variables in epidemiology. Int J Epidemiol **28,** 964–974.
45. Royston P, Sauerbrei W. (2005). Building multivariable regression models with continuous covariates in clinical epidemiology–with an emphasis on fractional polynomials. Methods Inf Med **44,** 561–571.
46. Sauerbrei W, Meier-Hirmer C, Benner A, Royston P. (2006). Multivariable regression building by using fractional polynomials: description of SAS, STATA and R programs. Comput Stat Data Anal **50,** 3646–3485.
47. Bates DM, Watts DG. (1988). Nonlinear Regression Analysis and its Applications. New York: Wiley.
48. Seber GAF, Wild CJ. (1989). Nonlinear Regression. New York: Wiley.
49. Ratkowsky DA. (1990). Handbook of Nonlinear Regression Models. New York: Marcel Dekker.
50. Liang K-Y, Zeger SL. (1986) Longitudinal data analysis using generalized linear models. Biometrika **73,** 13–22.
51. Burton P, Gurrin L, Sly P. (1998). Tutorial in biostatistics: extending the simple linear regression model to account for correlated responses: an introduction to generalized estimating equations and multi-level mixed modelling. Stat Med **17,** 1261–1291.
52. Hanley JA, Negassa A, Edwardes MD, Forrester JE. (2003). Statistical analysis of correlated data using generalized estimating equations: an orientation. Am J Epidemiol **157,** 364–375.
53. Brown H. (2006). Applied Mixed Models in Medicine, 2nd ed. Chichester, UK: Wiley.
54. McGilchrist CA. (1993). REML estimation for survival models with frailty. Biometrics **49,** 221–225.
55. Diez-Roux AV. (2000). Multilevel analysis in public health research. Annu Rev Public Health **21,** 171–192.
56. Little RJA, Rubin DB. (2002). Statistical Analysis with Missing Data, 2nd ed. Hoboken, NJ: Wiley.
57. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. (2006). Measurement Error in Nonlinear Models: A Modern Perspective, 2nd ed. London, UK: Chapman & Hall.

Chapter 10

Proteomics and Cancer Epidemiology

Mukesh Verma

Summary

Profiling of differentially expressed proteins is perhaps the most important and useful approach in developing tools for risk assessment in a population, diagnostic screening, and therapeutics. Proteomic markers have potential for identifying individuals at high risk of developing cancer; however, these markers have not been extensively used in cancer epidemiologic studies. Several markers have to be clinically validated. In this chapter, methods used in proteomic analysis of clinical samples, challenges in the proteomics and cancer epidemiology, and their potential solutions are discussed.

Key words: Artificial intelligence, 2-dimensional polyacrylamide gel electrophoresis (2D PAGE), biomarker, chromatin, epigenetics, histone, matrix-assisted Laser desorption ionization (MALDI) mass spectrometry, methylation, proteomics, risk assessment, surface-enhanced laser desorption ionization (SELDI) mass spectrometry, screening.

1. Introduction

Proteomics is rapidly becoming one of the most important tools for detection of the smallest possible changes in protein expression after the onset of a diseasem such as cancer. Proteomics refers to the study of the proteome. The term proteome was first coined in 1994, and it refers to all the proteins in a cell, tissue, or organism. Because proteins are involved in almost all biological activities, the proteome is a rich source of biological information. Proteomics represents an important medical advancement, and it may yield important tools in the war against cancer *(1,2)*. Although proteomics remains a paradigm shift from the genomic approach to medicine, it is important to remember that it is the protein that is the final arbiter of function, the levels at which molecular therapeutics are active. The term

Mukesh Verma (ed.), *Methods in Molecular Biology, Cancer Epidemiology, vol. 471*
© 2009 Humana Press, a part of Springer Science + Business Media, Totowa, NJ
Book doi: 10.1007/978-1-59745-416-2

proteomic approaches means the use of information from protein samples after their analysis for biomedical applications.

The human genome is estimated to contain approxmately 35,000 protein-encoding genes, and only a small percentage of proteins have been sequenced or identified *(2,3)*. The complete proteome has not been characterized for any organism to date. In contrast, the genome or the entire set of genes for several organisms has been sequenced, including in humans. Besides the difference in quantity, an important difference between the genome and proteome is that the genome is static and relatively unchanged from day to day. Cellular proteins, in contrast, are continually moving and undergoing changes, such as binding to a cell membrane; partnering with other proteins; gaining or losing a chemical group such as a sugar, fat, or phosphate; or breaking into two or more pieces. Proteins play a central role in the complex communication network within and between cells, and they are constantly responding to the needs of the organism. The protein analysis of biological samples has offered many opportunities and in identifying new tumor markers for screening, detection, diagnosis, treatment, and prognosis of cancer *(1,3–8)*. Proteomic approaches complement global gene expression approaches, which are powerful tools for identifying differentially expressed genes, but they are hampered by imperfect correlation of the levels of mRNA and protein *(9)*. To date, the most consistently successful proteomic methodology is the combination of two-dimensional polyacrylamide gel electrophoresis (2D PAGE) for protein identification, separation, and visualization, followed by mass spectrometric (MS) protein identification by using peptide mass fingerprints and tandem MS peptide sequencing. This technology has been discussed in detail. Methods and technologies currently used for protein analysis are discussed in this chapter. Potential implication of proteomic markers in cancer epidemiology and challenges in proteomics and cancer epidemiology also are discussed.

2. Selected Examples of Protein Markers in Samples from Cancer Patients

Most examples of protein markers discussed here are those that have been identified in biofluids. Proteins that serve as enzymes, hormones, structural elements, and antibodies have not been included. Emphasis is on multiplexing of markers and high throughput.

Differential protein expression profiles of biofluids detected by 2D PAGE and MS have been reported for various types of human cancers. Protein markers have been reported in individuals with bladder *(10,11)* breast *(12,13)*, colon *(14,15)*, esophagus

(16,17), gastric *(18)*, head and neck *(19,20)*, lung *(21)*, ovarian *(22)*, prostate *(23–28)*, skin *(29)*, and other cancer types *(30–43)* (**Tables 10.1** and **10.2**).

Some selected studies are described here, with a cautionary note that the sample size in all these studies was not sufficiently large to be called an ideal epidemiology study. These studies should be considered primitive examples only.

In the public database of the National Cancer Institute (www.cancer.gov), a population-based epidemiologic study has been proposed to examine proteomic detection and prediction of colon cancer. The study proposes to use a population-based nested case-control design, the unique resources of serial prediagnostic serum samples stored in a repository, collected epidemiologic information, and proteomic and bioinformatic technologies, a combination of which may make the study unique in addressing important scientific questions and overcoming major methodological limitations in serum proteomic research. The specific aims of the project are to 1) investigate whether there are proteomic profiles that discriminate between cases and controls; 2) investigate the earliest time point at which identified proteomic profiles occur, and assess whether proteomic profiles vary with time interval from the diagnosis; and 3) evaluate whether there are epidemiologic factors that may be related to the identified proteins. Cases will include approximately 400 colon cancer patients aged 17–79 years. Controls will be 400 individuals without a history of colon cancer of the same age, ethnicity, and time at the collection of each serum sample. Serum proteomics and glycoproteomics will be analyzed using the MALDI-TOF/TOF analyzer for serial serum samples from each study subject. Investigators also will measured serum insulin-like growth factor (IGF)-I and IGF-binding protein-3. Telephone interviews will be conducted to collect information on medical history, personal habits, occupational experiences, family history, reproductive history, and demographic characteristics. Information on pathologic diagnosis of colon cancer, medical history, occupational experiences, and other factors also is being extracted from the computerized health databases.

In a retrospective study, cases and controls were analyzed by immunophenotyping to follow up primary testicular lymphomas *(44)*. In renal cell carcinoma, proteomic levels of heat shock protein (HSP) 70 changed during progression of the disease, and they were used to follow up survival. In one gallbladder cancer patient survival study, p21(Wafl/Cip1) expression was found to be associated with longer survival if patients were positive for p53 mutation *(45)*. In a separate study, changes in prostate carcinoma incidence and manifestation during a 33-year period were investigated and included cases of prostate cancer diagnosed from 1972 to 2005 in Primorsko-Goranska County, Croatia *(46)*. The incidence rate was calculated in age-matched subjects. An 81.4-fold

Table 10.1
Proteomic markers used in cancer epidemiologic studies

Tumor type	Marker	Comment	Reference
Bladder cancer	X-linked inhibitor of apoptosis (XIAP)	Recurrence marker	*(10)*
Breast cancer	Mettaloproteinase-1	Screening and prognostic markers	*(87)*
Central nervous system	INI-1, few proteins in cerebrospinal fluid	Screening and prognostic marker	*(88,89)*
Endometrial cancer	PTEN, p53, beta-catenin	Correlation with survival was studied	*(47)*
Esophageal cancer	EGFR	Both phosphorylated and unphosphorylated forms were used for screening	*(90)*
Gall bladder cancer	p21	Patients with mutated p53 showed overexpression of p21 and a correlation with survival was observed	*(45)*
Gastric cancer	MMP-9, TIMP1	Prognostic markers	*(18)*
Liver cancer	N-Myc downregulated gene 1 (NDRG1)	Prognostic markers	*(91)*
Lung cancer	SOX, D52, CRMP1, MUC4	Small cell lung carcinoma	*(48,92,93)*
Ovarian cancer	CA125, p53, p21, p16, Mammaglobin B, MCAM	Over expression of p53 and p16 but low expression of p21 in ovarian cancer	*(94–98)*
Prostate cancer	PSA	Overexpression in prostate cancer	*(28, 46, 99)*
Renal cancer	HSP70, vascular cell adhesion molecule 1	Overexpression of nuclear and cytoplasmic protein in renal cancer progression; used as a prognostic marker and follow up survival	*(11,100)*
Skin cancer	Epidermal growth factor receptor, ezrin, hepatocyte growth factor, c-Met, bone morphogenetic protein 7	Prognostic marker	*(29,101)*
Urothelial cancer	Programmed death-1 (PD-1)/B7-H1 (also called PD-L1)	Prognostic marker	*(102)*

Table 10.2
Samples used for proteomic analysis

Sample	Cancer type
Cerebrospinal fluid	Brain
Nipple aspirate	Breast
Pancreatic fluid	Pancreatic cancer
Plasma and serum	Several cancers
Saliva and Sputum	Oral, head, and neck and lung cancer
Urine	Bladder cancer

incidence rate was observed from 1972 to 2005. Prostate-specific antigen (PSA) was analyzed in this study. The secondary cancer that developed in most of these patients was bone cancer.

One investigator followed cysteine levels in breast cancer epidemiologic study. Cysteine is the rate-limiting amino acid in biosynthesis of glutathione (GSH)-the key intracellular antioxidant and detoxifying agent. Recent research has suggested that high circulating levels of cysteine may prevent the development of breast cancer. Cysteinylglycine, a product of the extracellular GSH catabolism by γ-glutamyltranspeptidase (GGT1), is involved in the generation of reactive oxygen species. However, no study has assessed the relation of cysteinylglycine levels to breast cancer risk. Although genetic differences in GSH biosynthesis from cysteine and GSH catabolism could influence breast cancer risk and may modify the effect of cysteine and cysteinylglycine on risk of breast cancer, published reports are not there. In this prospective study of plasma cysteine, cysteinylglycine, gene variants involved in GSH biosynthesis from cysteine and GSH catabolism, and their interactive effects on risk of breast cancer in two large prospective cohorts, the Nurses' Health Study (NHS) and Nurses' Health Study II (NHSII) will be used. Investigators are evaluating common functional variants and determining the structure of haplotypes in three relevant candidate genes, including glutamate-cysteine ligase (also known as γ-glutamylcysteine synthetase, a rate-limiting enzyme catalyzing the first step of GSH biosynthesis from cysteine. with two subunits [GCLC, catalytic subunit; GCLM, regulatory subunit], glutathione synthetase (GSS, an enzyme catalyzing the second step of GSH biosynthesis), and GGT1 (an enzyme involved in the GSH catabolism to generate cysteinylglycine and in the transport of cysteine into the cell). Associations with these candidate genes may provide mechanistic evidence for the association between

plasma cysteine and cysteinylglycine and breast cancer risk. More than 2,000 incident breast cancer cases and their matched controls are being identified in the NHS and NHSII cohorts among women with achieved blood samples and free of cancer at baseline. The NHS and NHSII cohorts have many confirmed cases of breast cancer, long duration, high follow-up rates, and comprehensive covariate information, thus providing an extraordinary opportunity to examine the main effects of cysteine-related gene variants, plasma cysteine, and cysteinylglycine, and their interactions in relation to breast cancer risk. This study may help elucidate the role of cysteine in breast carcinogenesis, and it may suggest future approaches for breast cancer chemoprevention and suggest whether cysteine dietary supplements may improve cysteine status.

In endometrial cancer, prognostic significance of mutations of the phosphatase and tensin homolog deleted on chromosome ten (PTEN), p53, and β-catenin genes was investigated in a prospective study *(47)*. Levels of immunocytochemically expressed PTEN, p53, and β-catenin proteins in 80 fresh endometrial tumor specimens were determined. A correlation with survival also was noted. These proteins behaved differently as p53 staining correlated with increasing stage of lymph node metastases whereas β-catenin expression was significantly associated with decreased stage, and a negative lymph node status. PTEN positivity was correlated with decreased stage and negative lymph nodes. Furthermore, loss of PTEN also may be associated with a worse prognosis in patients with early-stage disease.

In squamous esophageal cancer, epidermal growth factor (EGFR) and its phosphorylated form (pEGFR) were used for diagnosis and prognosis. It is known that during cancer progression EGFR translocates from membrane to the nucleus. Specimens of esophageal squamous cell carcinoma (SCC) were examined immunohistochemically for their EGFR and pEGFR immunostaining patterns. Nuclear expression of pEGFR correlated with tumor, node, metastasis stage and lymph node metastasis, and it was associated with a poor outcome of esophageal SCC.

One cancer where protein biomarkers have been used successfully is lung cancer. Mucin 4 (MUC4) is a high-molecular-weight membrane-bound glycoprotein that is expressed in the foregut before epithelial differentiation. It is found in normal adult airway epithelium, non–small-cell lung carcinoma (NSCLC), and in other human malignancies independently of mucus secretion. MUC4 expression is common in pulmonary adenocarcinomas, and it may indicate a more favorable prognosis in early-stage adenocarcinomas. In a case-control study, immunohistochemical staining for MUC4 was performed on formalin-fixed, paraffin-embedded tissue samples from 343 cases of NSCLC arranged in

a high-density tissue microarray. These subjects were followed up to determine survival. MUC4 was frequently expressed in adenocarcinomas, squamous cell carcinomas, adenosquamous carcinomas, and large cell carcinomas. In patients with stage I and II adenocarcinoma, there was a trend toward longer patient survival with higher levels of MUC4 immunoreactivity compared with lower levels *(48)*.

3. Laboratory Methods in Proteomics

Methods described here are applicable to different cancers. 2D PAGE has been discussed in detail because this method has been used in maximum studies. Other methods have short descriptions.

Pancreatic cancer has been discussed in detail as an example because of its high incidence and high mortality rate in the United States *(49,50)*. Although not a common cancer, pancreatic cancer is the fourth leading cause of cancer death in the United States. There is an urgent need to identify molecular targets for early diagnosis and effective treatment of this devastating disease. Furthermore, selected proteomic markers that can distinguish cancer from noncancer cells during early stages of cancer development should be useful in identifying high-risk individuals in different populations with multiple ethnic groups.

3.1. Two-Dimensional Polyacrylamide Gel Electrophoresis

1. Cells are collected and lysed in buffer. One hundred fifty micrograms of protein is precipitated from the precipitation assay buffer using a Perfect-FOCUS kit (Geno Technology, Inc., St. Louis, MO).
2. Solubilization of the protein pellet in rehydration buffer, the first-dimension isoelectric focusing on a PROTEAN IEF cell, and the second-dimension sodium dodecyl sulfate (SDS)-PAGE separation using a precast Criterion 8 to 16% gradient gel in a Criterion Cell apparatus are performed according to the manufacturer's suggested protocols (Bio-Rad, Hercules, CA).
3. Protein extracts from each individual tissue sample are analyzed using immobilized pH gradient (IPG) strips with a pH range of 3.0 to 10.0.
4. Selected samples of each tissue type (i.e., normal, pancreatitis, and tumors) are pooled and then separated on IPG strips with pH ranges of 4 to 7 and 5 to 8. SDS-polyacrylamide gels are fixed, stained with SYPRO Ruby (Bio-Rad) overnight, and finally destained for 4 h.
5. Gel images are captured on a Kodak Image Station 440CF (Eastman Kodak, Rochester, NY).

3.2. Image Analysis of Two-Dimensional Gels

1. Integrated signal intensities are analyzed quantitatively using Kodak Image Station 440CF (Eastman Kodak Company, Rochester, NY).
2. Two independent observers then visually confirm differential expression.
3. The selected spots are manually excised and subjected to in-gel tryptic digestion based on the procedure described by Rosenfeld *et al. (51)*.

3.3. Protein Identification

1. The tryptic digests are analyzed on a Voyager-DE PRO MALDI-TOF MS (Applied Biosystems, Foster City, CA) as described previously *(52,53)*, with minor modifications.
2. The samples are desalted using a Ziptip μ-C18 pipette tip (Millipore Corporation, Billerica, MA).
3. The matrix is a saturated solution of α-cyano-4-hydroxycinnamic acid in 50% acetonitrile/0.1% trifluoroacetic acid diluted 1:1 with solvent.
4. For peptide mass mapping, automated MALDI-TOF spectral acquisition and database searching are performed.
5. Peptide mass lists from the mass range of 900 to 2,500 are filtered to remove trypsin autolysis, matrix, and keratin peaks.
6. The mass list of the 20 or 15 most intense monoisotopic peaks for each sample is entered in the search program MS-Fit 3.2.1 in the Protein Prospector suite *(54)*.
7. The National Center for Biotechnology Information nonredundant protein database is used for protein identification and analysis.
8. The search is performed for mammals using an 80-ppm peptide mass tolerance, and a minimum of six peptides matched required for a hit.
9. After intellical recalibration, a second search is performed with 30 ppm peptide mass tolerance and a minimum of five peptides matched.
10. The protein represented by the greatest number of peptide masses is matched, and the expected molecular weight is noted.
11. For samples with fewer than five peptides matched to one protein, at least one ion from the MS spectrum is subjected to manual postsource decay analysis for protein identification by peptide fragmentation.
12. The protein identified in each case is the highest scoring match, with the greatest number of ions matched to the peptide identified by the database search.

3.4. Western Blot Analysis

1. For one-dimensional immunoblotting, protein extracts (50 μg) are run on 8 to 16% gradient SDS-polyacrylamide gels, and the separated proteins are transferred onto a polyvinylidene diflouride membrane.
2. Antibodies specific to α-enolase, cathepsin D (Santa Cruz Biotechnology, Inc., Santa Cruz, CA), galectin-1 (Research Diagnostics, Inc., Flanders, NJ), calreticulin, copper zinc superoxide dismutase, and manganese superoxide dismutase (Nventa Biopharmaceuticals, Victoria, BC, Canada) are used at 1:500 dilutions.

3.5. Immunohistochemical Analysis

1. The antibodies used for immunohistochemistry are the same as those used for Western blot analysis.
2. The primary antibodies against α-enolase and cathepsin D are used at dilutions of 1:200, and antibodies against calreticulin and galectin-1 are used at dilutions of 1:400.
3. The biotinylated secondary antibody is used at a dilution of 1:200. The antibody complex is detected using ABC kit (Vector Laboratories, Burlingame, CA) with hematoxylin counterstain.
4. The primary antibody is replaced with phosphate-buffered saline as a negative control.

3.6. Surface-Enhanced Laser Desorption/Ionization (SELDI) MS Analysis

A quick and straightforward way to directly analyze laser Capture Microdissection (LCM)-purified samples is SELDI MS-TOF. Biofluids, such as blood sera, urine, and pancreatic juice, are good source of protein markers for cancer detection *(55–60)*. SELDI or MALDI with TOF MS detection, coupled with artificial intelligence-based informatics algorithms, have become a powerful tool with which to identify biomarker disease profiles.

Modern mass spectrometric instrumentation has achieved excellent performance in terms of sensitivity, resolution, and mass accuracy. Analysis of serum proteins can be achieved by SELDI MS. In one instance, sera from patients with American Joint Committee on Cancer stage I and II melanoma with negative loco-regional lymph nodes could identify potential melanoma-associated protein biomarkers of disease recurrence *(61)*. Protein patterns in breast cancer also have been studied by SELDI profile *(39,40,42)*.

The underlying principal of SELDI is surface-enhanced affinity capture through the use of protein chips consisting of chemical characteristics (hydrophobic, hydrophilic, cationic, and anionic) or biological ligands (proteins, receptors, antibodies, and DNA oligonucleotides). Sample is directly applied on the chip-surface, nonspecific binding partners are washed away, and the retained proteins are subsequently analyzed by TOF MS. SELDI has been successfully used to identify cancer-specific protein patterns in biofluids,

such as cerebral spinal fluid, seminal plasma, prostatic fluid, nipple aspirate, urine, saliva, and serum *(62–68)*. As with any new technology, reproducibility is critical. High-throughput assay robotics has been established, and the mass accuracy of SELDI spectra has reproducibility with coefficients of variation (CVs) of 0.5%. For those samples that were processed robotically, normalized intensity values for individual peaks are <20% CV. This is considered a well-accepted value in the proteomic field *(69–71)*. A modified Beckman BioMek2000 automatic robotic handling system (Beckam Coulter, Fullerton, CA) is used in clinical studies. At a time, 12 protein chips can be processed easily in this equipment. Reaction volume is kept 4–5% more to ensure that the robot has sufficient material to perform the processing, including compensation for pipette calibration problems and other related technical errors.

The discriminatory signature patterns that distinguish peaks between cancer and noncancer or other entities are different among scientific groups and report differing sensitivity and specificity. This may be due to the extent of the information cache used, e.g., up to 500,000 points (0–20 000 Da), and for that there may be more than a single reliable sensitive and specific diagnostic pattern. MS resolution capability varies between older and newer machines, proteomic chips containing different chromatographic surfaces, and the different bioinformatics analysis programs that reveal other discriminatory peaks. Advancement with this technology would occur with higher resolution instruments, positive identification of signature patterns by independent research groups, and identification of the molecules detected through MS that are the cancer diagnostic markers. It remains paramount that each profile put forward and each biomarker system undergoes rigorous blinded assessment using reliable validated standard operating procedures and cross-platform validation. To test the reproducibility of differentially expressed proteins, mass location and intensity from array to array on a single chip (intra-assay) and between chips (interassay), is achieved using a pooled normal serum quality control sample. Variation between spots within a chip is measured by Spearman correlation for the two quality control samples on each chip, and the median correlation is calculated *(72)*.

3.7. Artificial Intelligence

The first example where proteomics was used extensively to distinguish cancer and noncancer samples was ovarian cancer *(73–81)*. The discovery of protein patterns that distinguished ovarian cancer patients from those without disease relied on a sophisticated artificial intelligence computer program developed by Correlogic Systems, Inc. (Bethesda, MD). Scientists were able to "train" the computer to identify a pattern of only a handful of small proteins from thousands of candidates found in the blood that could distinguish cancer patients versus control samples. Once these

patterns were found, they were tested on other blinded samples from patients with and without cancer. Hundreds of samples were correctly identified using this approach. After completion of the Human Proteome Organization project, results from such studies can be compared against normal serum profile *(1)*. These results suggested that proteomic technology may help clinicians diagnose the disease much earlier than current setup. Advanced computational tools, data mining algorithms, and emerging data exchange format standards will facilitate the management, archiving, and electronic exchange and analysis of spectral data.

Several proteomic data formats are currently used for data exchange and data representation, including mzXML (XML format for MS) and the Human Proteome Organization data standards, such as mzData (XML format for MS), mzIdent (XML format for identification), and the Minimum Information About Proteomics Experiment. Standardized proteomics formats and algorithms can be incorporated into the software, and this would allow scientists to easily obtain and analyze proteomic data across various data repositories.

3.8. Laser Capture Microdissection

Tumor contains heterogeneous mixture of cells (e.g., stromal cells, inflammatory cells, extracellular matrix). For proteomic analysis, it is desirable to isolate pure cell population. LCM is such technique that facilitates the analysis of relatively pure cell population, but the procedure is very time-consuming and yields only small amount of material. For 2DE PAGE, approximately 50,000 cells need to be captured by the LCM. The time that is needed to obtain these cells influences the quality of the tissue slides and is a limiting factor for analyzing a large number of samples (especially in epidemiology projects). It is a good idea to use LCM for identification and validation of disease-specific protein profile and then to use standard methods for population screening. For SALDI or MALDI MS the number of cells needed is 200 to 1,000 *(62,82,83)*.

3.9. Isotope-Coded Affinity Tags (ICATs)

ICATs is a novel technology that facilitates quantitative proteomic analysis based on differential isotopic tagging of two pools of protein mixture that are being compared. This approach is based on the following components: isotope tagging of thiol-reactive group for the selective labeling of reduced cysteine (Cys) residues, and a biotin affinity tag to allow selective isolation of the labeled peptides. These two functional groups are joined by linkers that contain either eight hydrogen atoms (d0, light reagent) or eight deuterium atoms (d8, heavy reagent). Comparison between two specimens (e.g., mitochondria proteins form normal and disease cell) begins with the reduction of proteins from one pool that is then labeled with the isotopically light version of the ICAT reagent, whereas the other reduced pool of proteins is labeled by the isotopically heavy version of the ICAT reagent. Next, the two

samples are combined and digested with trypsin to generate peptide fragments. Cysteine-containing peptides are enriched by avidin affinity chromatography. There is an approximately 10-fold enrichment of the labeled peptides. The peptides may be further purified and analyzed by using a microcapillary reversed-phase liquid chromatography, followed by MS. The ratio of the isotopic molecular mass peaks that differ by 8 Da, provides a measure of the relative amounts of each protein from the original samples.

3.10. Protein Microarrays

Reverse-phase microarrays are used in clinical samples due to the ease with which they are generated, high reproducibility, built in dilution curves, and direct detection by using one antibody per microarray slide *(84–86)*. The technique has been used successfully for samples collected from breast *(86)* and lung cancer *(85)*. Protein microarrays can convey information regarding in vivo protein activation states for a variety of key signaling pathways. The technology is suitable for multiplexing because multiple samples can printed in tandem on the same slide, allowing quantitative comparisons of protein levels in different samples. In epidemiologic studies, sample lysates isolated from exposed and unexposed individuals may be printed on protein microarrays and analyzed.

4. Challenges and Potential Solutions

Challenges described in this section and potential solutions are applicable to proteomics field in general. Specific challenges for epidemiology field are difficult to analyze because properly designed studies with sufficient power where proteomic markers are used can be conducted only after proteomic markers have been validated in different studies. Although several proteomic biomarkers have been used in cancer studies, few challenges in the field still remain. The challenges of proteomic technology design and application arise from the complexity of protein dynamics and the vast amount of information required for characterizing the proteome. Innovative design and engineering strategies will be needed to overcome the current challenges facing proteomic research. Advanced technology improvements and developments can be made in sample preparation and storage, sample fractionation, automation, microfluidics, mass spectrometry, and informatics, and in affinity capture methods, including antibodies, aptamers, and protein arrays. Advanced data analysis methodologies and informatics algorithms are needed to address data preprocessing and management, peptide/protein identification, relative and direct quantification, elution time normalization, and peak alignment, signal processing, the extraction of diagnostic “molecular fingerprints,”

and peptide/protein sequencing within and among large data sets. This will require well-developed and characterized ontologies and bioinformatic annotations for use across the scientific community. The major challenges in the epidemiology studies are expensive tissue procurement, requirement of large number of samples, and false-positive and false-negative results. These challenges, however, are part of any epidemiologic study irrespective of the use of proteomic markers. One more challenge is related with environmental epidemiology. Exposure measurement errors in cancer epidemiology pose special methodologic challenges. For example, nutritional exposures form the basis for many etiologic hypotheses concerning cancer. However, nutrient intake is difficult to measure precisely. Other important examples considered in this area include genomic risk factors and environmental exposures. It is the role of measurement error correction methods to estimate the relationship between cancer outcomes and exposures. To accomplish this requires both a main study where disease and the surrogate exposure are measured and validation data to determine the extent of the measurement error. The major focus should be on nutritional studies based on intakes reported at a single survey when the target exposure is long-term average diet. In this case, the dependent variable is time to cancer incidence or mortality; thus, Cox models are considered in which time-varying covariates such as cumulative averages and cumulative exposures are of primary interest. Another approach is to extend methods for random and subject-specific error terms when the usual exposure measurements and the reference exposure measurements may be correlated.

5. Concluding Remarks

Comments in this section are applicable to proteomics field in general and not to epidemiology. Researchers are searching for proteins that can be used as early biomarkers of disease for risk assessment, or that may predict response to therapy or the likelihood of relapse after treatment in blood, urine, or diseased tissue. Despite tremendous advances in our understanding of the molecular basis of cancer, substantial gaps remain in our understanding of disease pathogenesis and in the development of effective strategies for early diagnosis and treatment. At this point, proteomics analyses are not mature enough to be used in the clinic as screening tools. However, these exploratory studies point to the promise of proteomics as a diagnostic marker. Validation in clinical trials in large groups of patients is necessary before proteomics patterns can be used routinely in the clinic as biomarkers for early disease. At the same time, techniques such as SELDI expression

profiling can be used to screens large number of samples because it is suitable for high-throughput analyses of small amount of material, such as body fluids, biospecimens, and microdissected cells, in a relatively short time. In cancer epidemiology, protein markers, in combination with genetic and epigenetic markers, should be used to identify risk factors and follow-up of the treatment. Generally, multiple biomarkers provide high sensitivity and specificity; therefore, when planning to use biomarkers in epidemiology studies, multiple markers should be included.

References

1. Hanash, S. (2004) HUPO initiatives relevant to clinical proteomics. *Mol Cell Proteomics* **3,** 298–301.
2. Omenn, G. S., States, D. J., Adamski, M., Blackwell, T. W., Menon, R., Hermjakob, H., Apweiler, R., Haab, B. B., Simpson, R. J., Eddes, J. S., Kapp, E. A., Moritz, R. L., Chan, D. W., Rai, A. J., Admon, A., Aebersold, R., Eng, J., Hancock, W. S., Hefta, S. A., Meyer, H., Paik, Y. K., Yoo, J. S., Ping, P., Pounds, J., Adkins, J., Qian, X., Wang, R., Wasinger, V., Wu, C. Y., Zhao, X., Zeng, R., Archakov, A., Tsugita, A., Beer, I., Pandey, A., Pisano, M., Andrews, P., Tammen, H., Speicher, D. W., and Hanash, S. M. (2005) Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics* **5,** 3226–3245.
3. States, D. J., Omenn, G. S., Blackwell, T. W., Fermin, D., Eng, J., Speicher, D. W., and Hanash, S. M. (2006) Challenges in deriving high-confidence protein identifications from data gathered by a HUPO plasma proteome collaborative study. *Nat Biotechnol* **24,** 333–338.
4. Imafuku, Y., Omenn, G. S., and Hanash, S. (2004) Proteomics approaches to identify tumor antigen directed autoantibodies as cancer biomarkers. *Dis Markers* **20,** 149–153.
5. Misek, D. E., Imafuku, Y., and Hanash, S. M. (2004) Application of proteomic technologies to tumor analysis. *Pharmacogenomics* **5,** 1129–1137.
6. Misek, D. E., Kuick, R., Wang, H., Galchev, V., Deng, B., Zhao, R., Tra, J., Pisano, M. R., Amunugama, R., Allen, D., Walker, A. K., Strahler, J. R., Andrews, P., Omenn, G. S., and Hanash, S. M. (2005) A wide range of protein isoforms in serum and plasma uncovered by a quantitative intact protein analysis system. *Proteomics* **5,** 3343–3352.
7. Song, K., and Hanash, S. (2006) Unraveling the complex proteome for biomarker discovery in gastrointestinal and liver diseases. *Gastroenterology* **131,** 1375–1378.
8. Steel, L. F., Haab, B. B., and Hanash, S. M. (2005) Methods of comparative proteomic profiling for disease diagnostics. *J Chromatogr* **815,** 275–284.
9. Xia, W. Y., Lien, H. C., Wang, S. C., Pan, Y., Sahin, A., Kuo, Y. H., Chang, K. J., Zhou, X., Wang, H., Yu, Z., Hortobagyi, G., Shi, D. R., and Hung, M. C. (2006) Expression of PEA3 and lack of correlation between PEA3 and HER-2/neu expression in breast cancer. *Breast Cancer Res Treat* **98,** 295–301.
10. Li, M., Song, T., Yin, Z. F., and Na, Y. Q. (2007) XIAP as a prognostic marker of early recurrence of nonmuscular invasive bladder cancer. *Chin Med J* **120,** 469–473.
11. Shioi, K., Komiya, A., Hattori, K., Huang, Y., Sano, F., Murakami, T., Nakaigawa, N., Kishida, T., Kubota, Y., Nagashima, Y., and Yao, M. (2006) Vascular cell adhesion molecule 1 predicts cancer-free survival in clear cell renal carcinoma patients. *Clin Cancer Res* **12,** 7339–7346.
12. Selvarajan, S., Wong, K. Y., Khoo, K. S., Bay, B. H., and Tan, P. H. (2006) Over-expression of c-erbB-2 correlates with nuclear morphometry and prognosis in breast carcinoma in Asian women. *Pathology* **38,** 528–533.
13. Short, S. M., Yoder, B. J., Tarr, S. M., Prescott, N. L., Laniauskas, S., Coleman, K. A., Downs-Kelly, E., Pettay, J. D., Choueiri, T. K., Crowe, J. P., Tubbs, R. R., Budd, T. G., and Hicks, D. G. (2007) The expression of the cytoskeletal focal adhesion protein paxillin in breast cancer correlates with HER2 overexpression and may help predict response to chemotherapy: a retrospective immunohistochemical study. *Breast J* **13,** 130–139.

14. Gosens, M. J., van Kempen, L. C., van de Velde, C. J., van Krieken, J. H., and Nagtegaal, I. D. (2007) Loss of membranous Ep-CAM in budding colorectal carcinoma cells. *Mod Pathol* **20,** 221–232.
15. Heinzerling, J. H., Anthony, T., Livingston, E. H., and Huerta, S. (2007) Predictors of distant metastasis and mortality in patients with stage II colorectal cancer. *Am Surg* **73,** 230–238.
16. Natsugoe, S., Uchikado, Y., Okumura, H., Matsumoto, M., Setoyama, T., Tamotsu, K., Kita, Y., Sakamoto, A., Owaki, T., Ishigami, S., and Aikou, T. (2007) Snail plays a key role in E-cadherin-preserved esophageal squamous cell carcinoma. *Oncol Rep* **17,** 517–523.
17. Sugiura, K., Ozawa, S., Kitagawa, Y., Ueda, M., and Kitajima, M. (2007) Co-expression of aFGF and FGFR-1 is predictive of a poor prognosis in patients with esophageal squamous cell carcinoma. *Oncol Rep* **17,** 557–564.
18. de Mingo, M., Moran, A., Sanchez-Pernaute, A., Iniesta, P., Diez-Valladares, L., Perez-Aguirre, E., de Juan, C., Garcia-Aranda, C., Diaz-Lopez, A., Garcia-Botella, A., Martin-Antona, E., Benito, M., Torres, A., and Balibrea, J. L. (2007) Expression of MMP-9 and TIMP-1 as prognostic markers in gastric carcinoma. *Hepato-gastroenterology* **54,** 315–319.
19. Smilek, P., Dusek, L., Vesely, K., Rottenberg, J., and Kostrica, R. (2006) Correlation of expression of Ki-67, EGFR, c-erbB-2, MMP-9, p53, bcl-2, CD34 and cell cycle analysis with survival in head and neck squamous cell cancer. *J Exp Clin Cancer Res* **25,** 549–555.
20. Ohshiro, K., Rosenthal, D. I., Koomen, J. M., Streckfus, C. F., Chambers, M., Kobayashi, R., and El-Naggar, A. K. (2007) Pre-analytic saliva processing affect proteomic results and biomarker screening of head and neck squamous carcinoma. *Int J Oncol* **30,** 743–749.
21. Zhong, L., Coe, S. P., Stromberg, A. J., Khattar, N. H., Jett, J. R., and Hirschowitz, E. A. (2006) Profiling tumor-associated antibodies for early detection of non-small cell lung cancer. *J Thorac Oncol* **1,** 513–519.
22. Materna, V., Surowiak, P., Markwitz, E., Spaczynski, M., Drag-Zalesinska, M., Zabel, M., and Lage, H. (2007) Expression of factors involved in regulation of DNA mismatch repair- and apoptosis pathways in ovarian cancer patients. *Oncol Rep* **17,** 505–516.
23. Fall, K., Garmo, H., Andren, O., Bill-Axelson, A., Adolfsson, J., Adami, H. O., Johansson, J. E., and Holmberg, L. (2007) Prostate-specific antigen levels as a predictor of lethal prostate cancer. *J Natl Cancer Inst* **99,** 526–532.
24. Garcia, J. A., Rosenberg, J. E., Weinberg, V., Scott, J., Frohlich, M., Park, J. W., and Small, E. J. (2007) Evaluation and significance of circulating epithelial cells in patients with hormone-refractory prostate cancer. *BJU Int* **99,** 519–524.
25. Gondi, V., Deutsch, I., Mansukhani, M., O'Toole, K. M., Shah, J. N., Schiff, P. B., Katz, A. E., Benson, M. C., Goluboff, E. T., and Ennis, R. D. (2007) Intermediate-risk localized prostate cancer in the PSA era: radiotherapeutic alternatives. *Urology* **69,** 541–546.
26. McFall, S. L. (2007) Use and awareness of prostate specific antigen tests and race/ethnicity. *J Urol* **177**, 1475–1480; discussion 1480.
27. Parekh, D. J., Ankerst, D. P., and Thompson, I. M. (2007) Prostate-specific antigen levels, prostate-specific antigen kinetics, and prostate cancer prognosis: a tocsin calling for prospective studies. *J Natl Cancer Inst* **99,** 496–497.
28. Pelzer, A. E., Bektic, J., Akkad, T., Ongarello, S., Schaefer, G., Schwentner, C., Frauscher, F., Bartsch, G., and Horninger, W. (2007) Under diagnosis and over diagnosis of prostate cancer in a screening population with serum PSA 2 to 10 ng/ml. *J Urol* **178,** 93–97; discussion 97.
29. Mallikarjuna, K., Pushparaj, V., Biswas, J., and Krishnakumar, S. (2007) Expression of epidermal growth factor receptor, ezrin, hepatocyte growth factor, and c-Met in uveal melanoma: an immunohistochemical study. *Curr Eye Res* **32,** 281–290.
30. Barry, W. T., Nobel, A. B., and Wright, F. A. (2005) Significance analysis of functional categories in gene expression studies: a structured permutation approach. *Bioinformatics (Oxford, England)* **21,** 1943–1949.
31. Czibere, A., Grall, F., and Aivado, M. (2006) Perspectives of proteomics in acute myeloid leukemia. *Expert Rev Anticancer Ther* **6,** 1663–1675.
32. Diamandis, E. P. (2004) Analysis of serum proteomic patterns for early cancer diagnosis: drawing attention to potential problems. *J Natl Cancer Inst* **96,** 353–356.
33. Fields, M. M., and Chevlen, E. (2006) Ovarian cancer screening: a look at the evidence. *Clin J Oncol Nurs* **10,** 77–81.

34. Gharib, T. G., Chen, G., Wang, H., Huang, C. C., Prescott, M. S., Shedden, K., Misek, D. E., Thomas, D. G., Giordano, T. J., Taylor, J. M., Kardia, S., Yee, J., Orringer, M. B., Hanash, S., and Beer, D. G. (2002) Proteomic analysis of cytokeratin isoforms uncovers association with survival in lung adenocarcinoma. *Neoplasia* **4,** 440–448.
35. Kolch, W., Mischak, H., and Pitt, A. R. (2005) The molecular make-up of a tumour: proteomics in cancer research. *Clin Sci (Lond)* **108,** 369–383.
36. Kumar, S., Mohan, A., and Guleria, R. (2006) Biomarkers in cancer screening, research and detection: present and future: a review. *Biomarkers* **11,** 385–405.
37. Lewis, S., and Menon, U. (2003) Screening for ovarian cancer. *Expert Rev Anticancer Ther* **3,** 55–62.
38. Preston, R. J. (2002) Quantitation of molecular endpoints for the dose-response component of cancer risk assessment. *Toxicol Pathol* **30,** 112–116.
39. Pusztai, L., Gregory, B. W., Baggerly, K. A., Peng, B., Koomen, J., Kuerer, H. M., Esteva, F. J., Symmans, W. F., Wagner, P., Hortobagyi, G. N., Laronga, C., Semmes, O. J., Wright, G. L., Jr., Drake, R. R., and Vlahou, A. (2004) Pharmacoproteomic analysis of prechemotherapy and postchemotherapy plasma samples from patients receiving neoadjuvant or adjuvant chemotherapy for breast carcinoma. *Cancer* **100,** 1814–1822.
40. Ricolleau, G., Charbonnel, C., Lode, L., Loussouarn, D., Joalland, M. P., Bogumil, R., Jourdain, S., Minvielle, S., Campone, M., Deporte-Fety, R., Campion, L., and Jezequel, P. (2006) Surface-enhanced laser desorption/ionization time of flight mass spectrometry protein profiling identifies ubiquitin and ferritin light chain as prognostic biomarkers in node-negative breast cancer tumors. *Proteomics* **6,** 1963–1975.
41. Scagliotti, G. V., and Novello, S. (2004) Current development of adjuvant treatment of non-small-cell lung cancer. *Clin Lung Cancer* **6(Suppl 2),** S63–70.
42. Seibert, V., Ebert, M. P., and Buschmann, T. (2005) Advances in clinical cancer proteomics: SELDI-ToF-mass spectrometry and biomarker discovery. *Brief Funct Genomics Proteomics* **4,** 16–26.
43. Verma, M. (2004) Biomarkers for risk assessment in molecular epidemiology of cancer. *Technol Cancer Res Treat* **3,** 505–514.
44. Al-Abbadi, M. A., Hattab, E. M., Tarawneh, M., Orazi, A., and Ulbright, T. M. (2007) Primary testicular and paratesticular lymphoma: a retrospective clinicopathologic study of 34 cases with emphasis on differential diagnosis. *Archives of pathology & laboratory medicine* **131,** 1040–1046.
45. Puhalla, H., Wrba, F., Kandioler, D., Lehnert, M., Huynh, A., Gruenberger, T., Tamandl, D., and Filipits, M. (2007) Expression of p21(Wafl/Cip1), p57(Kip2) and HER2/neu in patients with gallbladder cancer. *Anticancer Res* **27,** 1679–1684.
46. Spanjol, J., Maricic, A., Cicvaric, T., Valencic, M., Oguic, R., Tadin, T., Fuckar, D., and Bobinac, M. (2007) Epidemiology of prostate cancer in the mediterranean population of Croatia–a thirty-three year retrospective study. *Coll Antropol* **31,** 235–239.
47. Athanassiadou, P., Athanassiades, P., Grapsa, D., Gonidi, M., Athanassiadou, A. M., Stamati, P. N., and Patsouris, E. (2007) The prognostic value of PTEN, p53, and beta-catenin in endometrial carcinoma: a prospective immunocytochemical study. *Int J Gynecol Cancer* **17,** 697–704.
48. Kwon, K. Y., Ro, J. Y., Singhal, N., Killen, D. E., Sienko, A., Allen, T. C., Zander, D. S., Barrios, R., Haque, A., and Cagle, P. T. (2007) MUC4 expression in non-small cell lung carcinomas: relationship to tumor histology and patient survival. *Arch Pathol Lab Med* **131,** 593–598.
49. Verma, M. (2005) Pancreatic cancer epidemiology. *Technol Cancer Res Treat* **4,** 295–301.
50. Shen, J., Person, M. D., Zhu, J., Abbruzzese, J. L., and Li, D. (2004) Protein expression profiles in pancreatic adenocarcinoma compared with normal pancreatic tissue and tissue affected by pancreatitis as detected by two-dimensional gel electrophoresis and mass spectrometry. *Cancer Res* **64,** 9018–9026.
51. Rosenfeld, J., Capdevielle, J., Guillemot, J. C., and Ferrara, P. (1992) In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis. *Anal Biochem* **203**, 173–179.
52. Person, M. D., Lo, H. H., Towndrow, K. M., Jia, Z., Monks, T. J., and Lau, S. S. (2003) Comparative identification of prostanoid inducible proteins by LC-ESI-MS/MS and MALDI-TOF mass spectrometry. *Chem Res Toxicology* **16,** 757–767.
53. Jia, Z., Person, M. D., Dong, J., Shen, J., Hensley, S. C., Stevens, J. L., Monks, T. J., and Lau, S. S. (2004) Grp78 is essential for 11-deoxy-16,16-dimethyl PGE2-mediated cytoprotection in renal epithelial cells. *Am J Physiol* **287,** F1113–1122.

54. Clauser, K. R., Baker, P., and Burlingame, A. L. (1999) Role of accurate mass measurement (+/–10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Anal Chem* **71,** 2871–2882.
55. Negm, R. S., Verma, M., and Srivastava, S. (2002) The promise of biomarkers in cancer screening and detection. *Trends Mol Med* **8,** 288–293.
56. Srinivas, P. R., Verma, M., Zhao, Y., and Srivastava, S. (2002) Proteomics for cancer biomarker discovery. *Clin Chem* **48**, 1160–1169.
57. Verma, M., Kagan, J., Sidransky, D., and Srivastava, S. (2003) Proteomic analysis of cancer-cell mitochondria. *Nat Rev* **3,** 789–795.
58. Verma, M., and Srivastava, S. (2003) New cancer biomarkers deriving from NCI early detection research. *Rec Res Cancer Res* **163,** 72–84; discussion 264–266.
59. Verma, M., Wright, G. L., Jr., Hanash, S. M., Gopal-Srivastava, R., and Srivastava, S. (2001) Proteomic approaches within the NCI early detection research network for the discovery and identification of cancer biomarkers. *Ann N Y Acad Sci* **945,** 103–115.
60. Wagner, P. D., Verma, M., and Srivastava, S. (2004) Challenges for biomarkers in cancer detection. *Ann N Y Acad Sci* **1022,** 9–16.
61. Wilson, L. L., Tran, L., Morton, D. L., and Hoon, D. S. (2004) Detection of differentially expressed proteins in early-stage melanoma patients using SELDI-TOF mass spectrometry. *Ann N Y Acad Sci* **1022,** 317–322.
62. Espina, V., Wulfkuhle, J. D., Calvert, V. S., VanMeter, A., Zhou, W., Coukos, G., Geho, D. H., Petricoin, E. F., 3rd, and Liotta, L. A. (2006) Laser-capture microdissection. *Nat Protoc* **1,** 586–603.
63. Gillespie, J. W., Gannot, G., Tangrea, M. A., Ahram, M., Best, C. J., Bichsel, V. E., Petricoin, E. F., Emmert-Buck, M. R., and Chuaqui, R. F. (2004) Molecular profiling of cancer. *Toxicol Pathol* **32(Suppl 1),** 67–71.
64. Gulmann, C., Sheehan, K. M., Kay, E. W., Liotta, L. A., and Petricoin, E. F., 3rd (2006) Array-based proteomics: mapping of protein circuitries for diagnostics, prognostics, and therapy guidance in cancer. *J Pathol* **208,** 595–606.
65. Liotta, L. A., and Petricoin, E. F., 3rd (2003) The promise of proteomics. *Clin Adv Hematol Oncol* **1,** 460–462.
66. Petricoin, E. F., Belluco, C., Araujo, R. P., and Liotta, L. A. (2006) The blood peptidome: a higher dimension of information content for cancer biomarker discovery. *Nat Rev Cancer* **6,** 961–967.
67. Stone, J. H., Rajapakse, V. N., Hoffman, G. S., Specks, U., Merkel, P. A., Spiera, R. F., Davis, J. C., St Clair, E. W., McCune, J., Ross, S., Hitt, B. A., Veenstra, T. D., Conrads, T. P., Liotta, L. A., and Petricoin, E. F., 3rd (2005) A serum proteomic approach to gauging the state of remission in Wegener's granulomatosis. *Arthritis Rheum* **52,** 902–910.
68. Wulfkuhle, J. D., Edmiston, K. H., Liotta, L. A., and Petricoin, E. F., 3rd (2006) Technology insight: pharmacoproteomics for cancer--promises of patient-tailored medicine using protein microarrays. *Nat Clin Pract* **3,** 256–268.
69. Liggett, W. S., Barker, P. E., Semmes, O. J., and Cazares, L. H. (2004) Measurement reproducibility in the early stages of biomarker development. *Dis Markers* **20**, 295–307.
70. Schwegler, E. E., Cazares, L., Steel, L. F., Adam, B. L., Johnson, D. A., Semmes, O. J., Block, T. M., Marrero, J. A., and Drake, R. R. (2005) SELDI-TOF MS profiling of serum for detection of the progression of chronic hepatitis C to hepatocellular carcinoma. *Hepatology (Baltimore)* **41,** 634–642.
71. Semmes, O. J., Feng, Z., Adam, B. L., Banez, L. L., Bigbee, W. L., Campos, D., Cazares, L. H., Chan, D. W., Grizzle, W. E., Izbicka, E., Kagan, J., Malik, G., McLerran, D., Moul, J. W., Partin, A., Prasanna, P., Rosenzweig, J., Sokoll, L. J., Srivastava, S., Srivastava, S., Thompson, I., Welsh, M. J., White, N., Winget, M., Yasui, Y., Zhang, Z., and Zhu, L. (2005) Evaluation of serum protein profiling by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry for the detection of prostate cancer: I. Assessment of platform reproducibility. *Clin Chem* **51,** 102–112.
72. Kohli, M., Siegel, E., Bhattacharya, S., Khan, M. A., Shah, R., and Suva, L. J. (2006) Surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF MS) for determining prognosis in advanced stage hormone relapsing prostate cancer. *Cancer Biomark* **2,** 249–258.
73. Barbarini, N., Magni, P., and Bellazzi, R. (2006) A new approach for the analysis of mass spectrometry data for biomarker discovery. *AMIA Annu Symp Proc/AMIA Symp*, 26–30.

74. Kong, F., Nicole White, C., Xiao, X., Feng, Y., Xu, C., He, D., Zhang, Z., and Yu, Y. (2006) Using proteomic approaches to identify new biomarkers for detection and monitoring of ovarian cancer. *Gynecol Oncol* **100,** 247–253.
75. Kozak, K. R., Su, F., Whitelegge, J. P., Faull, K., Reddy, S., and Farias-Eisner, R. (2005) Characterization of serum biomarkers for detection of early stage ovarian cancer. *Proteomics* **5,** 4589–4596.
76. Lin, Y. W., Lin, C. Y., Lai, H. C., Chiou, J. Y., Chang, C. C., Yu, M. H., and Chu, T. Y. (2006) Plasma proteomic pattern as biomarkers for ovarian cancer. *Int J Gynecol Cancer* **16(Suppl 1),** 139–146.
77. Oh, J. H., Gao, J., Nandi, A., Gurnani, P., Knowles, L., and Schorge, J. (2005) Diagnosis of early relapse in ovarian cancer using serum proteomic profiling. *Genome Inform* **16,** 195–204.
78. Oh, J. H., Nandi, A., Gurnani, P., Knowles, L., Schorge, J., Rosenblatt, K. P., and Gao, J. X. (2006) Proteomic biomarker identification for diagnosis of early relapse in ovarian cancer. *J Bioinform Comput Biol* **4,** 1159–1179.
79. Wu, S. P., Lin, Y. W., Lai, H. C., Chu, T. Y., Kuo, Y. L., and Liu, H. S. (2006) SELDI-TOF MS profiling of plasma proteins in ovarian cancer. *Taiwan J Obstet Gynecol* **45,** 26–32.
80. Yu, J. S., Ongarello, S., Fiedler, R., Chen, X. W., Toffolo, G., Cobelli, C., and Trajanoski, Z. (2005) Ovarian cancer identification based on dimensionality reduction for high-throughput mass spectrometry data. *Bioinformatics (Oxford, England)* **21,** 2200–2209.
81. Zhang, H., Kong, B., Qu, X., Jia, L., Deng, B., and Yang, Q. (2006) Biomarker discovery for ovarian cancer using SELDI-TOF-MS. *Gynecol Oncol* **102,** 61–66.
82. Cowherd, S. M., Espina, V. A., Petricoin, E. F., 3rd, and Liotta, L. A. (2004) Proteomic analysis of human breast cancer tissue with laser-capture microdissection and reverse-phase protein microarrays. *Clin Breast Cancer* **5,** 385–392.
83. Jones, M. B., Krutzsch, H., Shu, H., Zhao, Y., Liotta, L. A., Kohn, E. C., and Petricoin, E. F., 3rd (2002) Proteomic analysis and identification of new biomarkers and therapeutic targets for invasive ovarian cancer. *Proteomics* **2,** 76–84.
84. Chatterjee, M., Ionan, A., Draghici, S., and Tainsky, M. A. (2006) Epitomics: global profiling of immune response to disease using protein microarrays. *Omics* **10,** 499–506.
85. Massion, P. P., and Caprioli, R. M. (2006) Proteomic strategies for the characterization and the early detection of lung cancer. *J Thorac Oncol* **1,** 1027–1039.
86. Vazquez-Martin, A., Colomer, R., and Menendez, J. A. (2007) Protein array technology to detect HER2 (erbB-2)-induced 'cytokine signature' in breast cancer. *Eur J Cancer* **43,** 1117–1124.
87. Lipton, A., Ali, S. M., Leitzel, K., Demers, L., Evans, D. B., Hamer, P., Brown-Shimer, S., Pierce, K., and Carney, W. (2007) Elevated plasma tissue inhibitor of metalloproteinase-1 level predicts decreased response and survival in metastatic breast cancer. *Cancer* **109,** 1933–1939.
88. Haberler, C., Laggner, U., Slavc, I., Czech, T., Ambros, I. M., Ambros, P. F., Budka, H., and Hainfellner, J. A. (2006) Immunohistochemical analysis of INI1 protein in malignant pediatric CNS tumors: lack of INI1 in atypical teratoid/rhabdoid tumors and in a fraction of primitive neuroectodermal tumors without rhabdoid phenotype. *Am J Surg Pathol* **30,** 1462–1468.
89. Khwaja, F. W., Reed, M. S., Olson, J. J., Schmotzer, B. J., Gillespie, G. Y., Guha, A., Groves, M. D., Kesari, S., Pohl, J., and Meir, E. G. (2007) Proteomic identification of biomarkers in the cerebrospinal fluid (CSF) of astrocytoma patients. *J Proteome Res* **6,** 559–570.
90. Hoshino, M., Fukui, H., Ono, Y., Sekikawa, A., Ichikawa, K., Tomita, S., Imai, Y., Imura, J., Hiraishi, H., and Fujimori, T. (2007) Nuclear expression of phosphorylated EGFR is associated with poor prognosis of patients with esophageal squamous cell carcinoma. *Pathobiology* **74,** 15–21.
91. Chua, M. S., Sun, H., Cheung, S. T., Mason, V., Higgins, J., Ross, D. T., Fan, S. T., and So, S. (2007) Overexpression of NDRG1 is an indicator of poor prognosis in hepatocellular carcinoma. *Mod Pathol* **20,** 76–83.
92. Ziv, T., Barnea, E., Segal, H., Sharon, R., Beer, I., and Admon, A. (2006) Comparative proteomics of small cell lung carcinoma. *Cancer Biomark* **2,** 219–234.
93. Tsutsumida, H., Nomoto, M., Goto, M., Kitajima, S., Kubota, I., Hirotsu, Y., Wakimoto, J., Hollingsworth, M. A., and Yonezawa, S. (2007) A micropapillary pattern is predictive of a poor prognosis in lung adenocarcinoma, and reduced surfactant apoprotein A expression in the micropapillary pattern is an excellent indicator of a poor prognosis. *Mod Pathol* **20,** 638–647.

94. Chauhan, S. C., Singh, A. P., Ruiz, F., Johansson, S. L., Jain, M., Smith, L. M., Moniaux, N., and Batra, S. K. (2006) Aberrant expression of MUC4 in ovarian carcinoma: diagnostic significance alone and in combination with MUC1 and MUC16 (CA125). *Mod Pathol* **19,** 1386–1394.
95. Buchynska, L. G., Nesina, I. P., Yurchenko, N. P., Bilyk, O. O., Grinkevych, V. N., and Svintitsky, V. S. (2007) Expression of p53, p21WAF1/CIP1, p16INK4A and Ki-67 proteins in serous ovarian tumors. *Exp Oncol* **29,** 49–53.
96. Tassi, R. A., Bignotti, E., Rossi, E., Falchetti, M., Donzelli, C., Calza, S., Ravaggi, A., Bandiera, E., Pecorelli, S., and Santin, A. D. (2007) Overexpression of mammaglobin B in epithelial ovarian carcinomas. *Gynecol Oncol* **105,** 578–585.
97. Petri, A. L., Hogdall, E., Christensen, I. J., Kjaer, S. K., Blaakaer, J., and Hogdall, C. K. (2006) Preoperative CA125 as a prognostic factor in stage I epithelial ovarian cancer. *Apmis* **114,** 359–363.
98. Aldovini, D., Demichelis, F., Doglioni, C., Di Vizio, D., Galligioni, E., Brugnara, S., Zeni, B., Griso, C., Pegoraro, C., Zannoni, M., Gariboldi, M., Balladore, E., Mezzanzanica, D., Canevari, S., and Barbareschi, M. (2006) M-CAM expression as marker of poor prognosis in epithelial ovarian cancer. *Int J Cancer* **119,** 1920–1926.
99. Thompson, I. M., and Ankerst, D. P. (2007) Prostate-specific antigen in the early detection of prostate cancer. *Cmaj* **176,** 1853–1858.
100. Ramp, U., Mahotka, C., Heikaus, S., Shibata, T., Grimm, M. O., Willers, R., and Gabbert, H. E. (2007) Expression of heat shock protein 70 in renal cell carcinoma and its relation to tumor progression and prognosis. *Histol Histopathol* **22,** 1099–1107.
101. Rothhammer, T., Wild, P. J., Meyer, S., Bataille, F., Pauer, A., Klinkhammer-Schalke, M., Hein, R., Hofstaedter, F., and Bosserhoff, A. K. (2007) Bone morphogenetic protein 7 (BMP7) expression is a potential novel prognostic marker for recurrence in patients with primary melanoma. *Cancer Biomark* **3,** 111–117.
102. Nakanishi, J., Wada, Y., Matsumoto, K., Azuma, M., Kikuchi, K., and Ueda, S. (2007) Overexpression of B7-H1 (PD-L1) significantly associates with tumor grade and postoperative prognosis in human urothelial cancers. *Cancer Immunol Immunother* **56,** 1173–1182.

Chapter 11

Different Study Designs in the Epidemiology of Cancer: Case-Control vs. Cohort Studies

Harminder Singh and Salaheddin M. Mahmud

Summary

It is only since the 1950s that most of the epidemiology studies on cancer have been conducted. The principal study designs for epidemiologic study of cancer etiology are case-control and cohort studies. These study designs have complimentary roles and distinct advantages and disadvantages. This chapter provides historical perspectives, description of the traditional and variant designs of case-control and cohort studies, and the relative advantages and disadvantages of these study designs.

Key words: Case control study, cancer etiology, neoplasia, screening.

1. Introduction

In the study of cancer epidemiology, clinical trials and experiments are often not feasible because it is unethical to expose individuals to a potential cause of disease simply to learn about the etiology of the disease. Moreover, the nonexperimental studies are often less expensive, and they can be performed over shorter intervals. The principal types of nonexperimental studies used to evaluate etiology of cancers are cohort studies and case-control studies. In cohort studies, groups with different exposures are followed up and compared to determine the differences in outcomes. In case-control studies, the initial step is sampling according to the outcomes in the study population to define the cases and controls; subsequently, the difference in rate of exposure among cases and controls is assessed.

Mukesh Verma (ed.), *Methods in Molecular Biology, Cancer Epidemiology, vol. 471*
© 2009 Humana Press, a part of Springer Science + Business Media, Totowa, NJ
Book doi: 10.1007/978-1-59745-416-2

2. Historical Perspective

Both case-control and cohort designs are relatively recent inventions. Case-control studies were rarely used in the first half of the 20th century; they came into mainstream use only in the later half of the century *(1)*. The development of the theoretical underpinning and the analytic methods were both motivated by challenges in understanding cancer etiology. The first known case-control study was probably a study of the role of reproductive experience in breast cancer undertaken by Lane-Claypon in England in 1926 *(2)*. Similar attempts in Britain (by Richard Doll and Bradford Hill) and in the United States (by Ernst Wynder and Evarts Graham) after World War II proved the effectiveness of case-control designs in establishing credible causal associations between environmental and lifestyle exposures and chronic illnesses, such as cancer and heart disease.

Over the years, case-control designs have become increasingly more acceptable as a valid alternative to the more costly cohort studies. This is the result of theoretical and technical advances in the statistical methods used to analyze data from case-control studies. First, Gerome Cornfield showed in a famous paper that the odds ratios calculated from case-control study are good approximations of rate ratios derived from an equivalent cohort study provided that the disease under study is rare (the famous rare disease assumption) *(3)*. Thus, exposure frequencies of cases and controls can be readily converted into a ratio of disease frequencies among the exposed and the unexposed. Then, in a paper that remains one of the most commonly cited papers in the biomedical literature, William Haenszel and Nathan Mantel introduced a procedure to adjust odds ratio for confounding by one or more variables *(4)*. Adjustment for confounding by multiple factors remained a difficult preposition until the introduction by Cornfield of logistic regression modeling of case-control data *(5)*. This method has been incorporated in many statistical software packages, and it is now the predominant analytical technique for case-control data *(1)*.

The first cohort studies also were concerned with establishing the etiologic link between smoking and lung cancer that was hinted at by the earlier case-control studies. At first called prospective studies by Richard Doll and Austin Bradford Hill in their famous 1954 paper that describes their "British Doctors Study," they were later renamed cohort studies by Brian MacMahon in 1960 *(6)*.

3. Cohort Studies in Cancer Research

3.1. Classic Cohort Studies

In classic cohort studies, two or more groups that are free of disease under study are identified. These groups referred to as

the study cohorts are different with regard to the exposure of interest. The study cohorts are followed-up to determine and compare the incidence (or mortality) rate of the disease of interest. The cohorts may be identified in present time and then followed-up prospectively over time (prospective cohort studies) or the cohorts may be identified at a specific time in the past (historical or retrospective cohort studies) and then followed-up to determine the disease occurrence. The advantage of prospective cohort studies is that one can plan and be rigorous in collecting the required information. In contrast, in historical cohort studies, one has to rely on the already collected information. The pay-off is in terms of time and expense, with historical cohort studies requiring less time and resources.

Closed cohort is a cohort with fixed membership. Once a closed cohort is defined no additional subjects can join the cohort. The closed cohorts gradually become smaller with size as individuals drop out due to mortality and loss of follow-up. Additional subjects can join open cohorts, which are also known as dynamic cohorts.

In the landmark prospective closed cohort study, the British Doctors Study, two thirds of male British doctors who responded to the questionnaire regarding their smoking habits in 1951 were identified *(7,8)*. These physicians were followed-up with periodic surveys, and their cause-specific mortality was determined from state records for the subsequent 50 yr, up to November 2001. Of the 34,439 initial respondents, 17% remained life-long nonsmokers. Smokers were found to have increased mortality from several disorders, including lung cancer, oropharyngeal cancer, ischemic heart disease, cerebrovascular disease, and chronic obstructive pulmonary disease.

3.2. "Linkage" Studies

With the advent of computers with superior computing power, large cohort studies have been performed with identification of retrospective cohorts from data routinely collected as part of the operation of health care systems. Exposures, longitudinal health service use, and outcomes can be determined by linkage of health use files and other databases that use key personal identifiers. To maintain patient confidentiality, data administrators usually provide linked data without unique identifiers, to the researchers.

As an example, we recently used Manitoba Health's physician billing claims database to identify individuals who had a negative colonoscopy (i.e., colonoscopic evaluation that did not result in a diagnosis of colorectal neoplasia) between April 1 1989 and December 31 2003 *(9)*. The cohort was then linked to the data in the provincial cancer registry and the population registry file so as to follow-up the cohort until diagnosis of colorectal cancer (CRC), death, or migration, whichever came first. Standardized incidence ratios were calculated to compare colorectal cancer incidence in our cohort with colorectal cancer incidence in the

provincial population. Stratified analysis was performed to determine the duration of the reduced risk of developing CRC. This was the first study to demonstrate that the risk of developing CRC remains decreased for >10 years after the performance of a negative colonoscopy.

3.3. New Advances: Repeated Measurement Studies

Repeated measurements designs are studies where exposure measurements, outcome measurements, or both are performed more than once for each participant. Examples include the following:

1. Pre-post design studies: For example, where an assessment such as blood pressure measurement is made before and after an intervention such as an administration of an antihypertensive medication. Pre-post designs are more commonly used in experimental studies or clinical trials.
2. Longitudinal studies: Typically two or more measurements are performed over time. These studies tend to be observational, and they could therefore be carried out prospectively or less commonly retrospectively using precollected data *(10)*. These studies are ideal for the study of complex phenomena such as the natural history of chronic diseases, including cancer. The repeated measurements and the longitudinal structure of the data allow the investigator to relate changes in time-dependent exposures to the dynamic status of the disease or condition under study. This is especially valuable if exposures are transient, and are not be measurable by the time the disease is detected.

As an example, Franco et al. *(11)* designed a repeated measurements longitudinal study to examine the natural history of cervical cancer as it relates to infections with the human papilloma virus (HPV). The investigators enrolled a systematic sample of 2,500 women from among attendees at a maternal and child health clinic in Sao Paulo, Brazil. These women were then followed up over a 10-year period in prescheduled returns every 4 months in the first year and twice yearly thereafter. Information on sociodemographic variables, ethnicity, smoking and drinking habits, reproductive history, and sexual practices was gathered at enrollment. At each subsequent visit, information on sexual and other behavioral factors since the preceding visit was collected using a supplementary questionnaire. At each visit, participants underwent a pelvic exam and had cervical specimens taken for Pap cytology, which permitted the detection of precursor lesions for cervical cancer, and HPV testing. A blood sample also was collected for serologic testing for HPV antibodies, micronutrients, and other biomakers.

Because of the frequent testing for cervical neoplasia and HPV infections, the investigators were able to measure the rates

of acquisition and clearance of cervical HPV infections and the rates of progression (and regression) of the various events that span the natural history of the disease from the initial HPV infection to the development of high-grade cervical neoplasia. Because information on sexual behavior, smoking, parity, and so on also was collected at each visit, it was possible to link changes in these behaviors to fluctuations in the risk of HPV infection, which allowed for deeper understanding of the complex interactions among host, environment, and viral factors.

This approach is well suited to the study of cervical carcinogenesis because of the availability of an ethical, safe, practical, and relatively cheap screening method (the Pap test), which allowed the researchers to periodically assess the cervical microenvironment. A similar approach to other cancers that lack acceptable screening tests (e.g., ovarian cancer) may not be possible. In fact, the lack of progress in understanding common cancers such as prostate cancer may reflect our inability to look back long enough and often enough to be able to detect transient or fluctuating but important risk factors.

The analysis of data from repeated measurement studies is far from straightforward especially when multiple outcomes per person are considered. The increased popularity of longitudinal studies with repeated measurements provided the impetus for the development of new statistical methods to account for the correlation between observations and to best use all the information collected by such studies. Generalized Estimating Equations models developed by Liang and Zeger at Johns Hopkins to extend generalized linear models to correlated data are increasingly being used in the analysis of these studies especially in relation to the study of human immunodeficiency virus/AIDS *(12,13)*.

4. Case-Control Studies in Cancer Research

4.1. Classic Design

In case-control studies, cases are identified from the source population and then their exposure status is evaluated. The incidence rates are not directly calculated because the exposure status of the entire population is not known; hence, the denominator for exposure cannot be determined. Instead, a control group is sampled from the source population to determine the relative size of exposed and unexposed denominators for the controls and cases, which in turn allows estimation of the relative incidence ratio (approximated using the odds ratio) *(14)*. In the landmark Wynder and Graham study, 51% of the 605 men with lung cancer reported prolonged heavy smoking (greater than a pack a day for the preceding 20 years) compared with 19% of the 780 controls

(15). The case-control studies are uniquely suited to study the etiology of cancers and other diseases of long induction period and rare occurrence.

4.2. Matched Case-Control Studies

Matched case-control studies involve selection of controls that are nearly identical to the cases with respect to one or more potential confounding factors. Individual matching is performed for a case at a time. Alternatively, frequency matching can be performed for an entire group of subjects; for example, for male cases, only males could be considered for selection as control subjects. Although matching does not prevent confounding, it does allow more efficient analysis to control for the confounding factors. Matching also can control for unknown confounders *(16)*. Alternative strategies for control of confounding, such as regression modeling, can result in biased results if there are inaccurate assumptions, such as presumed linear effect of the potential confounding variable.

Disadvantages of matching include a decrease in the ability to study the effect of the matching variables. Overmatching by factors that are related to the exposure of interest can result in cases and controls to be so similar that artificially no differences in the exposure of interest may be detected. Specific statistical analysis, which account for matching, are necessary.

4.3. Variants of Case-Control Design

4.3.1. Nested Case-Control Studies

In a nested case-control study, cases and controls are sampled from a preexisting and usually well-defined cohort. Typically, all subjects who develop the outcome under study during follow-up are included as cases. The investigator then randomly samples a subset of noncases (subjects who did not develop the outcome at the time of diagnosis of cases) as controls. Future cases can be selected as controls. The chance of a noncase being selected as a control is dependent upon the time spent by the noncase in the cohort. Nested case-control studies are more efficient than cohort studies because exposures are only measured for cases and a subset of noncases rather than all members of the cohort.

For example, to determine whether insulin-like growth factors (IGFs) play a role in the development of prostate cancer, Chan et al. *(17)* used a large cohort of 14,916 men who had been followed since 1982 as part of the Physicians' Health Study, a randomized, double-blinded, placebo-controlled trial of β-carotene and aspirin in U.S. male physicians. All men were free of cancer at enrollment. Before randomization, a blood sample was collected from each participant, which was then centrifuged. The resulting plasma was frozen and stored. Participants also filled out a self-administered questionnaire that included questions about relevant demographic, social, and medical risk factors. Using this information as well as participants' medical charts, the investigators

ascertained 152 cases of prostate cancer that occurred among the original cohort by 1992.

Chan et al. *(17)* carried out a nested case-control study by including all of the 152 prostate cancer patients as cases. For each case, a control was selected from among the men who were still alive and free of prostate cancer at the time of the case's diagnosis. IGF levels were then measured using the plasma that was collected at enrollment. They found that the mean level of IGF-1 among the cases (269.4 ng/ml) was significantly higher than among controls (248.9 ng/ml). Having an IGF-1 level >293 ng/ml was associated with a 4-fold increase in the risk of prostate cancer.

The nested case-control design is superior to a traditional case-control study because in the latter design, disease status (diagnosis of prostate cancer) and exposure status (levels of IGF-1) are measured at the same point in time, which complicates the interpretation of any detected associations. A traditional case-control design cannot rule out the possibility that prostate cancer is the cause, rather than a consequence, of raised IGF levels. Because nested designs typically use samples and data collected before the occurrence of the outcome, the direction (or temporality) of the association is less controversial. In the IGF study by Chan et al. *(17)*, an average of 7 years lapsed between the collection of plasma and the detection of prostate cancer making it less likely that elevated IGF levels were secondary to advanced prostate cancer.

Nested case-control designs share other strengths with prospective cohort studies. Again, because information is collected before the detection of the outcome and the determination of disease-control status, recall bias and certain types of observer bias are less likely. However, the main advantage of nested case-control designs is economic efficiency. Once the original cohort is established, nested case-control designs require far less financial resources and they are quicker in producing results than a new cohort. This is especially the case when outcomes are rare or when the measurement of the exposure is costly in time and resources. In the IGF study discussed here, determination of IGF levels and assortment of other biologic measurements were performed for 304 participants (all 152 cases and 152 randomly selected controls). The equivalent cohort analysis would have required these measurements to be made for all 14,916 participants! And what is more impressive is that those huge savings in time and resources come at very little loss in statistical power.

4.3.2. Case-Cohort Studies

Similar to the nested case-control study, in a case-cohort study, cases and controls are sampled from a preexisting and usually well-defined cohort. However, the selection of controls is not dependent upon the time spent by the noncases in the cohort.

Rather the adjustment for confounding effects of time is performed during analysis. Case-cohort studies are more efficient and economical than full cohort studies, for the same reasons as for nested case-control studies. In addition, the same set of controls can be used to study multiple outcomes or diseases *(18)*.

4.3.3. Case-Only Studies

For some hypotheses, the tenets of case-control methodology can be modified to produce specialized designs that are more efficient or less susceptible to bias and confounding than the classic case-control design.

For example, the case-only design is often used in genetic epidemiology to assess for the presence of interactions between genes and environmental exposures in disease causation. This design works because in many situations it is reasonable to assume that the distribution of a certain environmental factor (e.g., smoking) in the source population is independent from the possession of certain genotype (e.g., *N*-acetyltransferase haplotype). One can then study the distribution of smoking and *N*-acetyltransferase genotype among the cases of disease (e.g., breast cancer). Departure from the independence assumption among cases suggests the presence of interaction, i.e., the effect of smoking depends on the possession of certain *N*-acetyltransferase genotype.

5. Choice of Study Design

A major advantage of case-control studies is that multiple etiologies can be evaluated in the same study. Case-control and retrospective cohort studies are more efficient and less expensive than the prospective cohort or experimental studies. Cohort studies are more efficient to study exposures that are rare and responsible for only a small proportion of all cases of the disease under study.

Conversely, case-control studies allow only estimation of relative disease frequency; the actual incidence rate cannot be calculated. A major limitation of case-control studies is susceptibility to selection and recall bias; therefore, extra caution and planning are required to avoid these biases and ensure comparability between cases and controls. Selection of controls is perhaps the most important part in planning of case-control studies.

We conclude by emphasizing that case-control and cohort studies have complimentary roles in study of cancer epidemiology. When carefully performed, case-control studies and cohort studies can provide equivalent results.

References

1. Mahmud SM, Dean A, Franco EL. (2007). Epidemiological methods, a view from the Americas. In: Holland W, Olsen J, Florey C (eds.). The development of modern epidemiology. Oxford University Press, New York.
2. Breslow NE, Day NE. (1980). Statistical methods in cancer research. Volume I--The analysis of case-control studies. IARC Scientific Publications **32**, 1–350.
3. Cornfield, J. (1951). A method of estimating comparative rates from clinical data; applications to cancer of the lung, breast, and cervix. J Natl Cancer Inst **11(6),** 1269–1275.
4. Mantel, N, Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst **22(4),** 719–748.
5. Cornfield, J. (1962). Joint dependence of risk of coronary heart disease on serum cholesterol and systolic blood pressure: a discriminant function analysis. Fed Proc **21(4)Pt 2,** 58–61.
6. Doll R. (2001). Cohort studies: history of the method. I. Prospective cohort studies. Soz Praventivmed **46(2),** 75–86.
7. Doll R, Peto R, Boreham J, Sutherland I. (2004). Mortality in relation to smoking: 50 years' observations on male British doctors. BMJ **328(7455),** 1519.
8. Doll R, Hill AB. (2004). The mortality of doctors in relation to their smoking habits: a preliminary report. 1954. BMJ **328(7455),** 1529–1533.
9. Singh H, Turner D, Xue L, Targownik LE, Bernstein CN. (2006). Risk of developing colorectal cancer following a negative colonoscopy examination: evidence for a 10-year interval between colonoscopies. Jama **295(20),** 2366–2373.
10. Diggle PJ. (2002). The analysis of longitudinal data. Oxford University Press, New York.
11. Franco E, Villa L, Rohan T, Ferenczy A, Petzl-Erler M, Matlashewski G. (1999). Design and methods of the Ludwig-McGill longitudinal study of the natural history of human papillomavirus infection and cervical neoplasia in Brazil. Ludwig-McGill Study Group. Rev Panam Salud Publica **6(4),** 223–233.
12. Zeger SL, Liang KY. (1986). Longitudinal data analysis for discrete and continuous outcomes. Biometrics **42(1),** 121–130.
13. Zeger SL, Liang KY, Albert PS. (1988). Models for longitudinal data: a generalized estimating equation approach. Biometrics **44(4),** 1049–1060.
14. Rothman KJ, Greeenland S. (1998). Types of epidemiologic studies. In: Rothman KJ, Greeenland S (eds.). Modern epidemio-logy. Lippincott-Raven Publishers, Philadelphia, PA.
15. Wynder EL, Graham EA. (1985). Tobacco Smoking as a possible etiologic factor in bronchiogenic carcinoma. A study of six hundred and eighty-four proved cases. By Ernest L. Wynder and Evarts A. Graham. Jama **253(20),** 2986–2994.
16. Wacholder S, Silverman DT, McLaughlin JK, Mandel JS. (1992). Selection of controls in case-control studies. III. Design options. Am J Epidemiol **135(9),** 1042–1050.
17. Chan JM, Stampfer MJ, Giovannucci E, Gann PH, Ma J, Wilkinson P et al. (1998). Plasma insulin-like growth factor-I and prostate cancer risk: a prospective study. Science **279(5350),** 563–566.
18. Wacholder S, Silverman DT, McLaughlin JK, Mandel JS. (1992). Selection of controls in case-control studies. III. Design options. Am J Epidemiol **135(9),** 1042–1050.

Chapter 12

Methods and Approaches in Using Secondary Data Sources to Study Race and Ethnicity Factors

Sujha Subramanian

Summary

Race and ethnicity are increasing used in cancer research to assess differences in cancer incidence and response to therapy. In this chapter, we discuss the measurement and methodologic issues that should be addressed to minimize bias and derive valid estimates when performing such assessments. These issues include 1) lack of national standards for race and ethnicity categories; 2) difficulty in comparing race and ethnic categories in longitudinal assessments; 3) broad categorization of race and ethnicity groups that do not provide adequate details for meaningful assessments; 4) inaccuracies in race and ethnicity data collection, and 5) confounding by socioeconomics and other factors. Recommendations for improving race and ethnicity data collection also are discussed.

Key words: Race, ethnicity, methods, secondary data, cancer.

1. Introduction

Consistent variations have been observed in the incidence and mortality rates for cancer by race and ethnicity. Blacks have a higher incidence of certain cancers compared with whites. For example, black men and women have about twice the risk of their white counterparts in developing myeloma and stomach cancers *(1)*. In contrast, white males have 18 times higher risk of developing skin cancer than black males and four times the risk of developing testicular cancer. White females have double the risk of developing thyroid and brain cancers than black females. Differences also exist by ethnicity. For example, Hispanic women are twice as likely to be diagnosed with cervical cancer compared with non-Hispanic

Mukesh Verma (ed.), *Methods in Molecular Biology, Cancer Epidemiology, vol. 471*
© 2009 Humana Press, a part of Springer Science + Business Media, Totowa, NJ
Book doi: 10.1007/978-1-59745-416-2

women *(2)*. Even when census tract poverty rate is accounted for, most minority women have lower 5-year survival than non-Hispanic whites *(3)*.

There is considerable controversy over whether these disparities are indeed due to inherent biologic differences in race and ethnicity or whether they are merely manifestations of sociocultural factors *(4–8)*. Several genomic studies have demonstrated that although there are many common features among the races, there are indeed differences as well, and these differences can potentially influence the incidence and morality rates related to cancer *(9–11)*. For example, research indicates that the persistent findings of advanced state diagnosis among blacks and Hispanics compared with whites may be due to true biological differences that exist in breast cancer by race and ethnicity *(12–14)*. In contrast, there is a significant body of research that identifies the social, cultural, and economic factors that result in differences in cancer incidence and outcomes. The Institute of Medicine report ''Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care,'' synthesizes a large volume of literature describing racial and ethnic disparities in health care delivery *(15)*. Therefore, presently, there is evidence to support both these contrasting viewpoints and to highlight the complexity of the potential relationship between genetics and race *(6,11)*.

There is an unequal burden of cancer among the races; therefore, important information can be derived by assessing and explaining the variation in risk among the population subgroups. Such studies are required to not only evaluate the differential factors involved in etiology of cancer but also in diagnostic accuracy and treatment response. Secondary databases are playing an increasingly important role in epidemiology and public health research and in our understanding of racial differences *(16,17)*. Examples of secondary data sources include national surveys such as the National Health and Nutrition Examination Survey, claims data including those from the Medicare and Medicaid programs, and public vital statistics records such as death certificate data. Secondary data, especially linked data sets such as cancer registry data linked with administrative databases, provide a powerful tool to address research questions. With secondary data, unlike information gathered through primary data collection, categorizations and variable definitions have been predetermined and therefore present specific challenges to researchers. In this chapter, we discuss the measurement and methodological issues that should be addressed to minimize bias and derive valid estimates when performing assessments stratified by race and ethnicity. To provide a comprehensive review of the issues involved, both aspects directly related to secondary data analysis and those pertaining to race/ethnicity stratified assessments in general are presented.

2. Measurement and Methodological Issues

2.1. No National Standards for Race and Ethnicity Categories

The Office of Management and Budget (OMB) issued the Race and Ethnic Standards for Federal Statistics and Administrative Reporting contained in the Statistical Policy Directive No. 15 to provide minimum standard categories for reporting race and ethnicity. The description of these race categories are provided in **Table 12.1**. The OMB minimum standards include five racial categories (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, and

Table 12.1
Race and ethnicity categories included in the 1997 standards

Five race categories

American Indian or Alaska Native	A person having origins in any of the original peoples of North and South America (including Central America) and who maintains tribal affiliation or community attachment.
Asian	A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.
Black or African American	A person having origins in any of the black racial groups of Africa. Terms such as "Haitian" or "Negro" can be used in addition to "Black or African American."
Native Hawaiian or Other Pacific Islander	A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.
White	A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.
Ethnicity category	
Hispanic or Latino	A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race. The term "Spanish origin" can be used in addition to "Hispanic or Latino."

White) and two ethnic categories (Hispanic or Latino, and Not Hispanic or Latino). Despite this guidance, federal, state, and private institutions are not always coordinated in the coding and reporting of racial groupings, and this makes it difficult to compare results from these data sources *(18,19)*. The OMB standards have direct influence on federal data collection efforts and activities funded by federal dollars, which does not necessarily translate into national standards that are adopted for all race and ethnicity data collection. This lack of national standards makes comparisons across data sources and studies challenging as specific minority groups, such as Asians, may be assigned to different groupings *(11,20)*. Therefore, analysis or comparisons may have to be limited to those racial categories with common definitions across data sources and studies.

2.2. Comparability of Racial Categories in Longitudinal Assessments

OMB first issued the initial set of minimum standards in Statistical Policy Directive No. 15 in 1977. The standards in this directive were used for almost two decades throughout the federal government for record-keeping, collection, and presentation of data on race and ethnicity. Over time, the minimum categories defined by the standards came under increasing criticism for not reflecting the diversity of the U.S. population. In 1997, OMB issued revised standards that are currently in use today. The timeline for implementation and use of the race and ethnic data collection and reporting standards are chronicled in **Fig. 12.1**. Although the revised standards were introduced in 1997, during the interim period of several years, both the 1977 and 1997 standards could be used.

The 1997 "new" standards differed from the 1977 "old" standards in several ways, including the allocation of Pacific Islanders to their own category and not classified with Asians and identification by more than one racial category. Multiple race selection reported under the 1997 standards has to be allocated to a single race category to make them comparable with the race categories reported under the 1977 standards. "Bridge" methods can be used to compare or combine the data collected under the new derivative to the old derivative. A variety of bridge methods have been proposed *(19)*. The simplest of these approaches is to exclude all those who report multiple races from the analysis. Other methods include whole allocation to either the largest group or the smallest group. If a person selected both Black and American Indian, then under the largest group allocation method the person would be assigned to "Black" and under the smallest group allocation method to "American Indian." More complex methods involve fractional allocations to race groups. The assignment can be based on equal fractions to each race group reported. Another approach would be to use fractional assignments based on surveys that have asked people who report

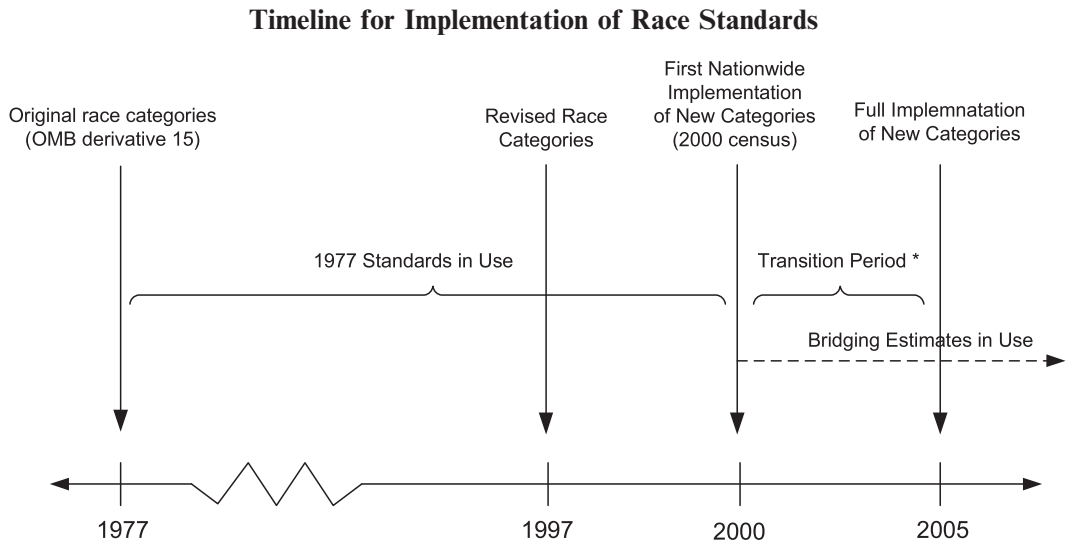


Fig. 12.1. Timeline for implementation of race standards.

more than one race which race they would select if only one race category was available. The National Center for Health Statistics and other researchers also have used regression-based estimates to develop complex algorithms for collapsing multiple races into single race categories *(21,22)*. Each of these bridge methods have their advantages and disadvantages, but in general they provide estimates that are not substantially different *(4)*. Evaluations of whole assignment and fractional assignment methods indicate that fractional assignment provides a closer approximation to past racial distributions *(4,23–25)*. The racial categories that were most sensitive to choice of bridging method were those with numerically small populations with a high proportion of multiracial respondents, that is, the American Indian and Alaska Native category, and to a lesser extent, the Asian category *(4)*.

Although the number of multiracial individuals is presently small, this number is expected to grow substantially *(4,26,27)*; therefore, it is important to identify and adopt a bridging method that provides unbiased estimates. Selection of appropriate bridging methodology is important for cancer research because trends in incidence and mortality are often reported, and pooled samples are required to analyze minority populations *(28,29)*.

2.3. Broad Categorization of Race and Ethnicity Groups

The level of detail collected on racial and ethnic groupings is not always sufficient to perform in-depth assessments. Studies have shown that there is substantial variation within racial subgroups, and this variation is often even more pronounced than that between broadly specified groups recommended by the OMB Directive 15 *(19)*. Pidena et al. *(30)*, for example, have shown that Japanese women have significantly better survival when

diagnosed with breast cancer than other Asian women. Pooled analysis of all Asian women will therefore not be able to identify these within group differences. Similarly, the incidence of colorectal cancer differs among subgroups with origins in the Middle East *(31)* who are often categorized as white *(32)*. The currently specified minimum race categories therefore may not be detailed enough to provide the necessary information for in-depth assessment of racial differences.

An additional issue is that in some cases (e.g., 2000 census), mixed race individuals only select a category to indicate "more than one race." This response does not provide any information on specific race, which significantly limits the information available for analysis. Even when such data are available, an additional challenge that researchers face is the availability of adequate sample sizes to perform assessments stratified by subgroups. Research also has shown that some racial groups are more likely to be mixed. For example, African Americans come from a broad range of African ancestry; therefore, additional details are required to reclassify them into coherent groups. Also, studies using admixed populations (e.g., African Americans) generally require careful assessment to avoid biases. Traditional statistical adjustment and matching may not adequately control for potential biases, and researchers have to use additional approaches to limit bias *(33–36)*.

2.4. Quality of Data on Race and Ethnicity

Research has shown that asking individuals to identify their own race or ethnicity, that is, self-reported information, results in the most reliable data *(37)*. Several studies have assessed self-reported race and ethnicity with information available from administrative data. Medicare enrollment data, for example, is fairly accurate in identifying whites (97%) and blacks (95%) but far less so for other races or Hispanic origin (<60%) *(38, 40)*. Comparison of race reported in Veterans Administration database and self report indicate poor agreement between the data sources, with least agreement for Native American, Asian, and Pacific Islander patients *(41)*. These low rates of agreement (approx.. 60%) were largely the result of including administrative data from patients whose race/ethnicity was unknown in the comparisons.

Several studies have assessed the accuracy of race information available in cancer surveillance databases *(42–47)*. Registry data are reviewed and certified by the North American Association of Central Cancer Registries, and these data standards include predetermined formats for the collection of both racial and ethnic data. Cancer registries, though, obtain data on race, Hispanic ethnicity, and birthplace primarily from hospital records, and these assignments may involve subjective appraisals by hospital personnel *(48)*. One study that compared self-reported race and

ethnicity information with the data available in the cancer registry found that the accuracy for non-Hispanic whites and blacks was high (>90%), moderate for Hispanics and some Asian subgroups (70–90%), and very low for American Indians (<20%). Overall, the disagreement between registry and self-reported race/ethnicity was 11% *(49)*. A recent study that compared Surveillance, Epidemiology and End-Stage Results data with self-reported data from the National Longitudinal Mortality Study also reported similar results; excellent agreement for white race, moderate for Hispanic ethnicity, and low for American Indians/Alaska Natives *(50)*. Another study that compared racial classifications in cancer registry data with self-report found that ethnic classification by medical record alone resulted in an underestimate of Hispanic cancer cases and incidence rates *(43)*. This bias was reduced when medical records and surnames were used together to classify Hispanic cancer cases. In addition, there is evidence of variation in the collection of race and ethnicity information *(49)*. Hospitals that provide information on cancer cases do not consistently collect information on race and ethnicity, and they generally do not have established policies regarding the collection of these data.

2.5. Relationship between Race/Ethnicity and Socioeconomic Factors

The interrelationship between race/ethnicity and socioeconomics has been well studied, and there are decades of evidence that at least some of the racial and ethnic patterns of cancer risk are due to socioeconomic differences *(51–55)*. Studies have found that the incidence rate of cancer can vary as much, if not more, by socioeconomic status than by race *(56)*. In addition, the direction of the relationship also differs when specific types of cancers are analyzed. For example, breast cancer incidence increases with affluence among Hispanic women, and incidence of cervical cancer increases as socioeconomic status decreases among all four racial/ethnic groups, with the strongest association present for white women *(55)*. This interrelationship is seen both among adults and the pediatric population. A recent study by Peace et al. *(57)* showed that there were differences in cancer incidence between Hispanic children living in Florida and California due to familial, socioeconomic, or environmental factors.

This complex relationship between race/ethnicity and socioeconomics results in confounding when assessing the impact of race on cancer incidence and treatment outcomes. In **Fig. 12.2**, the impact of this interrelationship along the continuum of cancer care is illustrated. This confounding can be addressed through matching of the cohorts or restrictions of the samples *(58)*. The most common approach to address the confounding of race and socioeconomic status is to use multivariate methods *(55)*, including multiple regression analysis with propensity scores. Regression techniques require a sufficient sample of individuals

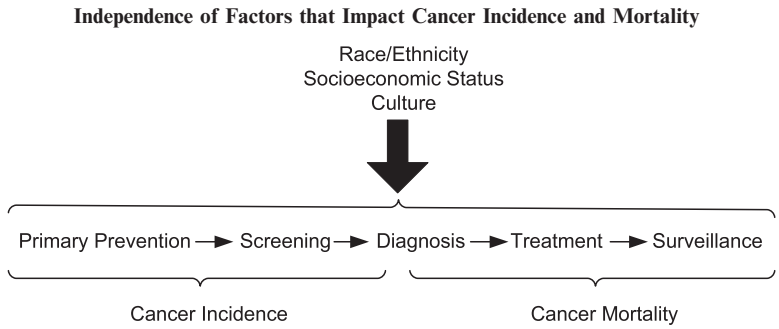


Fig. 12.2. Independence of factors that impact cancer incidence and mortality.

in each of the stratifications to provide reliable estimates. Population based samples generally do not have large enough number of minorities to provide sufficient statistical power for analysis. Therefore, stratified sampling to ensure adequate sample of comparative racial/socioeconomic status groups is essential.

3. Discussion

Some researchers have argued against using race in health analysis because of inaccuracies in measurement and instead suggest that other life factors, such as culture and socioeconomic status that are causally related to health outcomes, should be used *(59)*. Race has shown in many instances to independently impact cancer incidence and outcomes even after controlling for socioeconomics and other factors *(3,60)*. Therefore, race and ethnicity are useful variables *(61,62)* that researchers will continue to use in cancer research. Analysis stratified by race and ethnicity helps us understand both genetic and environmental influence on cancer incidence and response to therapy.

As outlined in this chapter, there are several measurement and methodologic issues that should be addressed in designing, analyzing, and reporting studies on differences by race and ethnicity. It is critical that a thorough assessment of the potential limitations and direction of bias are performed in these studies to draw attention to the possible inaccuracies in race and ethnicity measurement. Race and ethnicity are essential stratifications in health disparities research, and there is increasing recognition that although current national data are useful to study these issues there remains significant gaps that need to be addressed in the future. The Panel on the Department of Health and Human Services' Collection of Race and Ethnicity Data *(63)* and other experts have outlined several recommendations to improve data

collection on race and ethnicity including the following. First, common definitions should be adopted to help standardize race and ethnicity data collection. Second, these measures should be collected in all health and health care data systems, and third, the systems that do collect these data should improve accuracy of the data collected (e.g., Medicare administrative data). Fourth, appropriate sampling techniques should be used to ensure that adequate sample of minority groups are available to perform meaningful analysis. Improvement in the collection of race and ethnicity data will ensure that future research can result in targeted interventions that can reduce the burden of cancer for all racial and ethnic groups.

References

1. American Cancer Society. (2007b). Cancer facts and figures for African Americans 2007–2008. American Cancer Society, Atlanta, GA.
2. Morbidity and Mortality Weekly Report. (2002). Invasive cervical cancer among Hispanic and non-Hispanic women–United States, 1992–1999. *Morbid Mortal Wkly Rep* **51(47)**, 1067–1070.
3. Ward, E., Jemal, A., Cokkinides, V., et al. (2004). Cancer Disparities by Race/Ethnicity and Socioeconomic Status *CA Cancer J Clin* **54(2),** 78–93.
4. Lee, S. M. (2001). Using the new racial categories in the 2000 census. Washington, DC: The Annie E. Casey Foundation and The Population Reference Bureau.
5. Braun, L. (2002). Race, ethnicity, and health: can genetics explain disparities? *Perspect Biol Med* **45,** 159–174.
6. Sankar, P., Cho, M., Condit, C., et al. (2004). Genetic research and cancer disparities. *J Am Med Assoc* **291,** 2985–2989.
7. Burchard, E., Ziv, E., Coyle, N., Gomez, S., Tang, H., Karter, A., et al. (2003). The importance of race and ethnic background in biomedical research and clinical practice. *N Engl J Med* **348(12),** 1170–1175.
8. Cooper, R. S. (2003). Race, genes, and health—new wine in old bottles? *Int J Epidemiol* **32,** 26–28.
9. Brown, L., Hoover, R., Greenberg, R., et al. (1994). Are racial differences in squamous cell esophageal cancer explained by alcohol and tobacco use? *J Natl Cancer Instit* **86,** 340–1345.
10. Zahm, S. H., and Fraumeni, J. F. J. (1995). Racial, ethnic, and gender variations in cancer risk: considerations for future epidemiologic research. *Environ Health Perspect* **103(Suppl 8),** 283–286.
11. Rebbeck, T. R., Halbert, C. H., and Sankar, P. (2006). Genetics, epidemiology, and cancer disparities: is it black and white? *J Clin Oncol* 24 **(14),** 2164–2169.
12. Watlington, A., Byers, T., Mouchawar, J., et al. (2007). Does having insurance affect differences in clinical presentation between hispanic and non-hispanic white women with breast cancer? *Cance,* **109(10),** 2093–2099.
13. Velanovich, V., Yood, M., Bawle, U., et al. (1999). Racial difference in the presentation and surgical management of breast cancer. *Surgery* **125(4),** 375–379.
14. Gapstur, S., Dupuis J, Gann P, et al. (1996). Hormone receptor status of breast tumors in black, hispanic, and non-hispanic white women. an analysis of 13,239 cases. *Cancer* 77**(8),** 1465–1471.
15. Geiger, H. (2003). Racial and ethnic disparities in diagnosis and treatment: a review of the evidence and a consideration of causes. In: Unequal treatment: confronting racial and ethnic disparities in healthcare. Washington, DC: National Academies Press.
16. Smigal, C., Jemal, A., Ward, E., et al. (2006). Trends in breast cancer by race and ethnicity: update 2006. *CA Cancer J Clin* **56(3),** 168–183.
17. Kadan- Lottick, N., Ness, K., Bhatia, S., et al. (2003). Survival variability by race and ethnicity in childhood acute lymphoblastic

leukemia. *J Am Med Assoc* **290(15),** 2008–2014.

18. HHS Data Council Working Group on Racial and Ethnic Data. (1999). Improving the collection and use of racial and ethnic data in HHS. HHS Data Council Working Group on Racial and Ethnic Data. http://aspe.hhs.gov/datacncl/racerpt/. Cited 6 Jun 2007.
19. Friedman, D. J., Cohen, B. B., Averbach, A. R., and Norton, J. M. (2000). Race/ethnicity and OMB Directive 15: implications for state public health practice. *Am J Public Health* **90(11),** 1714–1719.
20. Hirschhorn, J., Lohmueller, K., Byrne, E., et al. (2002). A comprehensive review of genetic association studies. *Genet Med* 4, 45–61.
21. Parker, J. D., Schenker, N., Ingram, D. D., et al. (2004). Bridging between two standards for collecting information on race and ethnicity: an application to census 2000 and vital rates. Volume 119: Public Health Reports.
22. Liebler, C. A.,and Halpern- Manners, A. (2006). A practical approach to using multiple-race response data: a bridging method for public-use microdata. Department of Sociology and Minnesota Population Center, University of Minnesota, Minneapolis, MN.
23. Grieco, E. M. (2002). An evaluation of bridging methods using race data from census 2000. *Popul Res Policy Rev* **21,** 91–107.
24. Parker, J. D. and Makuc, D. M. (2002). Methodologic implications of allocating multiple-race data to single-race categories. *Health Serv Res* **37(1),** 201–213.
25. Heck, K., Parker, J., McKendry, C., and Chavez, G. (2003). Mind the gap: bridge methods to allocate multiple-race mothers in trend analyses of birth certificate data. *Matern Child Health J* **7(1),** 65–70.
26. Hirschman, Charles, Richard Alba, and Reynolds Farley. "The Meaning and Measurement of Race in the U.S. Census: Glimpses into the Future." *Demography* 37:381–393, 2000.
27. Waters, M. C. (2000). Immigration, intermarriage, and the challenges of measuring racial/ethnic identities. *Am J Public Health* **90(11),** 1735–1737.
28. Puukka, E., Stehr- Green, P., and Becker, T. M. (2005). Measuring the health status gap for American Indians/Alaska Natives: getting closer to the truth. *Am J Public Health* **95(5),** 838–843.
29. Boscoe FP, Ward MH, Reynolds P. Current practices in spatial analysis of cancer data: data characteristics and data sources for geographic studies of cancer. Int J Health Geogr. 2004 Dec 1;3(1):28.
30. Pineda, M. D., White, E., Kristal, A. R., and Taylor, V. (2001). Asian breast cancer survival in the U.S.: a comparison between Asian immigrants, US-born Asian Americans and caucasians. *Int J Epidemiol* **30,** 976–982.
31. Moshkowitz, M., and Arber, N. (2005). Differences in incidence and distribution of colorectal cancer among races and ethnic societies: lifestyle, genes or both? *Digestion* **72,** 219–222.
32. Hasnain- Wynia, R., and Baker, D. W. (2006). Obtaining data on patient race, ethnicity, and primary language in health care organizations: current challenges and proposed solutions. *Health Serv Res* **41(4),** 1501–1518.
33. Yanga, Q., Rabinowitzc, D., Isasib, C., and Sheac, S. (2000). Adjusting for confounding due to population admixture when estimating the effect of candidate genes on quantitative traits. *Hum Hered J* **50,** 227–233.
34. Wang, Y., Localio, R., and Rebbeck, T. (2004). Evaluating bias due to population stratification in case-control association studies of admixed populations. *Genet Epidemiol* **27,** 14-20.
35. Wang, Y., Localio, R., and Rebbeck, T. (2005). Consideration of the minimal number of unlinked markers required to correct for bias due to population stratification in candidate gene association studies. *Hum Hered J* **59,** 165–175.
36. Khlat, M., Cazes, M.-H., Genine, E., et al. (2004). Robustness of case-control studies of genetic factors to population stratification: magnitude of bias and type error. *Cancer Epidemiol Biomarkers Prev* **13,** 1660–1664.
37. Boehmer, U., Kressin, N. R., Dan R., et al. (2002). Self-reported vs administrative race/ethnicity data and study results. *Am J Public Health* **92(9),** 1471-1473.
38. Waldo, D. (2004). Accuracy and bias of race/ethnicity codes in the Medicare Enrollment Database. *Health Care Financing Rev* **26,** 61–72.
39. Arday, S., Arday, D., Monroe, S., and Zhang, J. (2000). HCFA's racial and ethnic data: current accuracy and recent improvements. *Health Care Financing Rev* **21(4),** 107–116.
40. McAlpine DD, Beebe TJ, Davern M, Call KT. Agreement between self-reported and administrative race and ethnicity data among Medicaid enrollees in Minnesota.

Health Serv Res. 2007 Dec; 42(6 Pt 2): 2373–88.

41. Kressin, N. R., Chang, B.-H., Hendricks, A., and Kazis, L. E. (2003). Agreement between administrative data and patients' self-reports of race/ethnicity. *Am J Public Health* **93(10),** 1734–1739.
42. Stewart, S., Swallen, K., Glaser, S., Horn-Ross, P., and West, D. (1998). Adjustment of cancer incindence rates for ethnic misclassification. *Biometrics* **54,** 774–781.
43. Stewart, S. L., Swallen, K. C., Glaser, S. L., Horn- Ross, P. L., and West, D. W. (1999). Comparison of methods for classifying hispanic ethnicity in a population-based cancer registry. *Am J Epidemiol* **149(11),** 63–71.
44. Swallen, K., West, D., Stewart, S., Glaser, S., and Horn- Ross, P. (1997). Predictors of misclassification of Hispanic ethnicity in a population-based cancer registry. *Ann Epidemiol* **7,** 200–206.
45. Swallen, K., Glaser, S., Stewart, S., West, D., Jenkins, C., and McPhee, S. (1998). Accuracy of racial classification of vietnamese patients in a population-based cancer registry. *Ethn Dis* **8,** 218–227.
46. Lin, S., O'Malley, C., and Lui, S. (2001). Factors associated with missing birthplace information in a population based cancer registry. *Ethn Dis* **11,** 598–605.
47. Lin, S., Clark, C., O'Malley, C., and Le, G. (2002). Birthplace and survival among Asian women diagnosed with breast cancer in cancer registry data: the impact of selection bias [letter]. *Int J Epidemiol* **12,** 511–513.
48. Blustein, J. (1994). The reliability of racial classifications in hospital discharge abstract data. *Am J Public Health* **84,** 1018—1021.
49. Gomez, S., and Glaser, S. (2006). Misclassification of race/ethnicity in a population-based cancer registry (United States) *Cancer Causes Control* **17(6),** 771–781.
50. Clegg LX, Reichman ME, Hankey BF, Miller BA, Lin YD, Johnson NJ, Schwartz SM, Bernstein L, Chen VW, Goodman MT, Gomez SL, Graff JJ, Lynch CF, Lin CC, Edwards BK. Quality of race, Hispanic ethnicity, and immigrant status in population-based cancer registry data: implications for health disparity studies. Cancer Causes Control. 2007 Mar;18(2):177–87. Epub 2007 Jan 11.
51. Devesa, S., and Diamond, E. (1980). Association of breast cancer and cervical cancer incidences with income and education among whites and blacks. *J Natl Cancer Inst* **65,** 515–528.
52. Devesa, S., and Diamond, E. (1983). Socioeconomic and racial differences in lung cancer incidence. *Am J Epidemiol* **118,** 818–831.
53. McWhorter, W., Schatzkin, A., Horm, J., and Brown, C. (1989). Contribution of socioeconomic status to black/white differences in cancer incidence. *Cancer* **63,** 982–987.
54. Williams, D. R. (1996). Race/ethnicity and socioeconomic status: measurement and methodological issues. *Int J Health Serv* **26,** 483–505.
55. La Veist, T. (2005). Disentangling race and socioeconomic status: a key to understanding health inequalities. *J Urban Health* **82(2 Suppl 3),** 26–34.
56. Krieger, N., Quesenberry, C. J., Peng, T., Horn- Ross, P., et al. (1999). Social class, race/ethnicity, and incidence of breast, cervix, colon, lung, and prostate cancer among Asian, Black, Hispanic, and White residents of the San Francisco Bay Area, 1988–92 (United States). *Cancer Causes Control* **10(6),** 525–537.
57. Peace, S., Button, J., Trapido, E., Mac Kinnon, J., et al. (2005). Cancer incidence among hispanic children in the United States. *Pan Am J Public Health* **18(1),** 5–13.
58. Karter, A. J. (2003). Commentary: race, genetics, and disease—in search of a middle ground. *Int J Epidemiol* **32,** 26–28.
59. Fullilove, M. (1998). Abandoning "race" as a variable in public health research–an idea whose time has come. *Am J Public Health* **88,** 1297–1298.
60. Singh, G. K., Miller, B., Hankey, B., and Edwards, B. (2004). Persistent area socioeconomic disparities in U.S. incidence of cervical cancer, mortality, stage, and survival, 1975–2000. *Cancer* **101(5),** 1051–1057.
61. Buehler, J. (1999). Abandoning race as a variable in public health research. *Am J Public Health* **89(5),** 783–784.
62. Rabin, R. (199). The use of race as a variable in public health research. *Am J Public Health* **89**(5), 783.
63. National Research Council. (2004). Eliminating health disparities: measurement and data needs: panel on dhhs collection of race and ethnicity data. Committee on National Statistics-Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Chapter 13

Statistical Methods in Cancer Epidemiological Studies

Xiaonan Xue and Donald R. Hoover

Summary

In this chapter, we discuss statistical methods for various study designs that are commonly used in epidemiological research and particularly in cancer epidemiological research. After a brief review of basic concepts in epidemiological studies, statistical methods for case-control studies and cohort studies are discussed. Statistical methods for nested case-control and case-cohort studies, which have been increasingly used in cancer epidemiology, also are discussed. This chapter is designed for cancer epidemiologists who understand basic statistical methods for commonly used epidemiological study designs and are able to initiate power and sample size calculations. Therefore, this chapter emphasizes newly developed statistical methods for epidemiological studies as well as study planning.

Key words: Case-control study, matched case-control study, cohort study, nested case-control study, case-cohort study.

1. Introduction

This chapter is designed for cancer epidemiologists who understand basic statistical methods for commonly used epidemiological study designs and are able to initiate power and sample size calculations. This chapter therefore emphasizes newly developed statistical methods for epidemiological studies as well as study planning. In **Subheading 1**, we review basic concepts in epidemiological studies. We then discuss statistical methods for case-control studies in **Subheading 2** and cohort studies in **Subheading 3**. In **Subheading 4**, we introduce statistical methods for nested case-control and case-cohort studies. Due to space limitations, the methods are briefly described, but references for further details on each statistical method are given. Although approaches for power and sample size estimation are discussed,

Mukesh Verma (ed.), *Methods in Molecular Biology, Cancer Epidemiology, vol. 471*
© 2009 Humana Press, a part of Springer Science + Business Media, Totowa, NJ
Book doi: 10.1007/978-1-59745-416-2

due to space limitations the focus is on explaining how to implement available software as opposed to theoretical explanation.

1.1. Risk Measures and Simple Adjustment for Age

Common disease risk measures include prevalence and incidence probabilities or rates. Measures of exposure and disease association, defined based upon comparison of prevalence and incidence between two or more groups, include risk difference (attributable risk), relative risk (risk ratio), odds ratio (OR), and hazards ratio. However, direct comparison of rates between two groups may be subject to a variety of distortion (confounding) perhaps most commonly due to age differences between groups. For example, if two exposure groups do not have the same age (or socioeconomical status, sex, etc.) distribution and age (or socioeconomic status, sex, etc.) is associated with the outcome, the differences in age (or socioeconomic status, sex, etc.) can distort the observed association between exposure and outcome. Thus, it is often useful to standardize rates before comparison, in particular to a common age distribution. The principle behind direct standardization is to calculate hypothetical rates for each comparison group using an identical age (or socioeconomic status, sex, etc.) distribution. This identical age distribution is denoted as the standard. A different type of standardization, often referred to as indirect standardization, is to apply a standard set of age-specific rates weighted to the age distribution of the population under study. The comparison between the study population and the standard is often presented as a standardized mortality ratio (SMR), which is ratio of an observed number of cases in the study population to an expected number if the study population had the same rate of disease as the standard.

1.2. Population and Samples

A population is the collection of all individuals of interest. However, population risk measures are often estimated from a subset study sample as it is unfeasible to measure all individuals in large populations. Based on sample estimates, one hopes to draw conclusions about the population from which the sample was selected. Therefore, it is essential for the sample to be representative of the study population. Most of the methods that we cover herein assume that random samples are used. A random sample is one selected in a way that gives each member of the population an equal probability to be chosen. There are two desirable features of a random sample: first, it eliminates bias; and second, it enables one to determine the reliability of the estimates.

Commonly used sampling designs include cross-sectional studies, case-control studies, and cohort studies. We discuss these designs in detail herein.

1.3. Statistical Inference from Samples

Statistical inference from samples includes point and interval estimation of the population parameter(s) of interest, such as proportion with disease, OR, and hazard ratio; and hypothesis

testing. The point estimate, that is, the best guess of the true population parameter based on the data from the sample is often obtained using the maximum likelihood estimation method. The maximum likelihood estimator has an asymptotic normal distribution. Confidence interval limits for the population parameter can thus be generated based on normality. Statistical software is commonly used to obtain these estimators. In this chapter, we provide guidance on how to use SAS (SAS Institute, Cary, NC), a commonly used and powerful statistical package to perform the analysis.

When a hypothesis test is being conducted, two types of errors can be made: type I error is the probability of mistakenly rejecting the null hypothesis, and type II error is the probability of mistakenly rejecting an alternative hypothesis. When multiple hypotheses are being tested in the same study, unless these tests are hypothesis-driven rather than exploratory, the type I error can be inflated therefore the issue of multiple comparisons need to addressed. A Bonferroni adjustment is often used to do this. The idea is to reduce the type I error for each hypothesis test so that the global type I error of falsely rejecting any of the null hypotheses when all are true does not exceed a certain limit. However, a Bonferroni test is often too conservative. Benjamin and Hochberg *(1)* proposed a method to control the false discovery rate: the proportion of wrong findings among all the tests that are statistically significant. When a large number of multiple hypotheses are tested at the same time, for example, when a large scale of single nucleotide polymorphisms (SNPs) are being examined in association with the development of colon cancer, even Benjamin and Hochberg's approach can be too conservative. Other methods such as the permutation-based test *(2)* and the two-stage approach *(3)* should be applied instead.

1.4. Sample Size and Power

One important aspect in study designs is to determine an adequate sample size. For random samples, the primary source of error for the point estimate is random variation. Such variation depends on the heterogeneity of individuals in the population and the size of the sample. The smaller the heterogeneity and the larger the sample size, the more accurate the point estimate. The sample size can then be determined by the accuracy needed for the point estimate.

Often, an epidemiological study is designed to answer a scientific question, which translates to a statistical hypothesis. The adequacy of the sample size is determined by the power of the study, that is, the probability of rejecting the null hypothesis when a given alternative hypothesis is true. Usually at least 80% power is considered to be adequate. The sample size required to achieve a certain power typically for the comparison of two groups depends on the following: 1) the within-group individual

variation, 2) magnitude of the likely or minimally important group differences one needs to see, and 3) the size of the type I error. The computation of sample size often also should take into consideration other factors such as loss of data (including loss of follow-up and missing data), measurement error, and misclassification.

Sample size/power formulas have been developed for most of the study designs described in **Subheading 2 and Subheading 3**, and software is available to implement these formulas so that, in general, we do not have to manually calculate the power or sample sizes. For example, PASS 2003 *(4)* is a commonly used program to calculate sample size or power for a variety of study designs and statistical analyses. In this chapter, we use PASS to illustrate the computation for power and sample size.

2. Design and Analysis of Case-Control Studies

2.1. Case-Control Studies

2.1.1. Introduction to Cross-Sectional and Case-Control Studies

In a cross-sectional study, a random group of subjects are sampled at a given time. The association between a prior or current exposure and a current condition or outcome is evaluated. But exposure assessment may be subject to recall and measurement bias. It is difficult for cross-sectional studies to obtain enough persons to evaluate associations involving rare outcomes or exposures. Also, this design is useless for outcomes such as death, which removes persons from a population. Therefore, case-control studies are more commonly used.

In a case-control study, a group of individuals with the outcome (cases) and a comparable (sometimes matched) group without the outcome (controls) are identified; previous exposures to risk factor for cases and controls are assessed *(5)*. For example, 500 patients (cases) who died of liver cancer and 1000 controls who did not are identified and age matched (on the control being alive at the age the case died). The cases and the controls are compared on lifetime smoking (before the age of death for the case) to see whether it is associated with increased risk of liver cancer death. The case-control study is less expensive and time-consuming than is a cohort study. Although, unlike a cohort, in a case control study the assessment of exposure is still subject to recall bias, it is more easily to obtain enough events (cases) for a rare outcome than will a cross-sectional study.

2.1.2. Unadjusted and Simple Adjusted Methods on Exposure-Disease Association

In case-control studies, the strength of the association between exposure and disease is typically measured by an OR. Consider a simple case of a binary exposure, the OR is defined as the ratio between the odds of exposure given the presence of the

Table 13.1
Distribution of exposure for cases and controls

	Exposed	Unexposed	Total
Case	a	b	n_1
Control	c	d	n_0

disease (case) and the odds of exposure given the absence of the disease (control). Suppose n_1 cases and n_0 controls are selected. The distribution of exposure for cases and controls is given in Table 13.1.

Let the true population OR be denoted by , it is estimated from the table by

$$\hat{\psi} = \frac{ad}{bc}$$

with $\hat{\text{var}}(\ln \hat{\psi}) = \left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}\right)$. Therefore, a $(1-\alpha)$ confidence interval for Ψ can be estimated by

$$\left(\hat{\psi} \,/\, \exp\left(z_{1-\alpha/2}\sqrt{\hat{\text{var}}(\ln \hat{\psi})}\right), \hat{\psi} \exp\left(z_{1-\alpha/2}\sqrt{\hat{\text{var}}(\ln \hat{\psi})}\right)\right).$$

The OR also can be interpreted as the ratio between the odds of disease given exposure and the odds of disease given nonexposure. When the disease is rare (<10%), the OR is a good approximation of relative risk of disease between exposure and nonexposure; that is, p_e/p_u where p_e is the risk of disease in an exposed person and p_u is the risk of the disease in an unexposed person. If none of the cells have small numbers (i.e., "expected frequency" less than 5) *(6)*, a hypothesis test of association between exposure and disease can be computed using a chi-square test or by checking whether the confidence interval for the OR specified above includes 1; if some cells have small numbers, the test of association can be better obtained by Fisher's exact test or by checking if the exact confidence interval for the OR includes 1 *(7)*. When there are more than two levels of exposure and there is a natural order in the levels of the exposure variable, Armitage *(8)* and Mantel *(9)* developed tests for detecting alternative hypotheses of a trend in disease risk with increasing levels of exposure.

As noted, the existence of confounding, if not properly controlled, can distort the apparent association between exposure and disease. One method to control confounding is to divide the sample into a series of strata that are each internally homogeneous with respect to the confounding factor(s). Separate ORs can be calculated within each stratum and be tested for equality *(7)*.

The Mantel and Haenszel (M-H) test *(10)* can be used to test for global null hypotheses on the adjusted (within stratum) ORs. An underlying common pooled OR across all levels of the confounding factor can be estimated using the M-H estimate *(10)*. Several methods have been proposed to estimate the confidence interval for the common OR *(6)*. The method proposed by Robins et al. *(11)* is more satisfactory among them since it is more robust to the size and the number of the strata.

When there is more than one confounding factor, each exposure category is defined as a particular combination of factor levels. Then, the same methods described above can be used to evaluate the association between exposure and disease with cross-stratification. However, computations can be burdensome because so many strata must be created. In other cases, the confounders could be continuous with a dose-response association that does not lend itself well to strata. Furthermore, some of confounding factors may act as an effect modifier, in which case the assumption of a common OR is not valid. Therefore, in this setting, multivariate analyses are best carried out using regression models described in **Subheading 2.1.3**.

2.1.3. Multivariate Analysis: Logistic Regressions

Logistic regression models can be applied to make inference on exposure–disease association for case-control studies while adjusting for multiple and/or continuous covariates. In a case-control design, the case and control statuses are fixed, and the dependent variable is the exposure variable of interest. However, it is not difficult algebraically to manipulate the case-control likelihood function to obtain a logistic regression model in which the dependent variable is the disease status. Let Y denote the case status, Y = 1 is a case and Y = 0 is a control; X denote the primary exposure variable and W_1 W_2, …. W_K be K potential confounding factors. Define p as the probability to be a case in this study and logit(p) = log(p/1 – p)). Then a logistic regression model to estimate the exposure–disease association is defined as the follows:

$$\text{Logit } p = \beta_0 + \beta_{1\chi} + \gamma_1 w_1 + \gamma_2 w_2 + \ldots + \gamma_K w_K \qquad (2.1)$$

where e^{β_0} is not interpretable unless the sampling probabilities for cases and controls are known, and e^{β_1} is the OR for disease between the exposed and the nonexposed while controlling for W_1 W_2, …. W_K. The exposure variable also can be continuous, the interpretation of e^{β_1} is then the OR for disease associated with per unit increase in X. For $\beta_1 = 0$, $e^{\beta_1} = 1$ meaning no association between the exposure and outcome adjusting for the W's. The logistic regression model is estimated by maximizing the log-likelihood function; thus, the estimate of $\beta_1 : \hat{\beta}_1$ is the maximum likelihood estimate with asymptotic normality properties. The variance of $\hat{\beta}_1$ is estimated by the inverse of the

information matrix of the log-likelihood function. The $(1-\alpha)$ confidence interval for e^{β_1} is $\exp\left(\hat{\beta}_1 \pm z_{1-\alpha/2}\sqrt{\hat{\text{var}}(\hat{\beta}_1)}\right)$ The test for null hypothesis at a given type 1 error α can be based on a score test on β_1 or by a likelihood ratio test, which is equivalent to seeing if the $(1-\alpha)$ confidence interval for e^{β_1} includes 1. Detailed discussion on modeling fitting, checking, and statistical inferences can be found in Agresti *(12)* and Hosmer and Lemeshow *(13)*.

Model 1.1 implies that the OR associated with exposure is the same at each level of W_k. Homogeneity of OR is only true if none of the W_k's is an effect modifier. To test for homogeneity of the OR across different levels of W (for simplicity consider W_1), an interaction term between X and W_1 is included in the model as the product of X and W_1:

$$\text{Logit } p = \beta_0 + \beta_{1\chi} + \gamma_1 w_1 + \delta x w_1 + \gamma_2 \mathbf{w}_2 + \ldots + \gamma_K w_K$$

For example, if sex is included in the model as a potential effect modifier with two levels: male = 0 and female = 1, then the OR associated with exposure is given in **Table 13.2**. Therefore, e^{δ} is the ratio of the ORs between exposure and outcome for females versus males. The test of homogeneity in OR between age groups is thus equivalent of testing $\delta = 0$. Again, the estimate $\hat{\delta}$ and its variance are obtained using maximum likelihood estimates. Logistic regression can be implemented using PROC LOGISTIC and PROC GENMOD in SAS *(14,15)*.

2.1.4. Sample Size and Power Estimation

Estimation of sample size and power for case-control studies are usually based on the comparison of cases and controls for level of exposure.

2.1.4.1. Binary Exposures Unadjusted

When exposure is a binary variable (yes, no) and no confounding exists, then using notation from PASS power and sample size depends on combinations of the following parameters: the type 1 error α; the portion of controls exposed P1; the portion of cases exposed (under the expected alternative) P2; the number of controls N1; the number of cases N2 and the type II error β (not to be confused with the β used in equation 2.1).

Example 2.1 If

Table 13.2
Effect modification by Gender

Gender	OR for exposure
Male	e^{β_1}
Female	$e^{\beta_1+\delta}$

- Based on prior information P1 = 40% of controls and P2 = 60% of cases are expected to be exposed (which corresponds to an OR of 2.25);
- A two-sided hypothesis test with $\alpha = 0.05$ will be used;
- Equal numbers of cases and controls will be used, that is, allocation ratio = 1.0, then as calculated in PASS (**Table 13.3**), to achieve 80% power to detect a significant association between exposure and disease, N1 = 97 controls and N2 = 97 cases are needed.

The number of controls does not have to be the same as the number of cases. The standard deviation of the estimate of ln(OR) is a function of $\sqrt{1+\frac{1}{m}}$, where m is the number of controls per case. Therefore, when controls are "cheap" compared with cases, having more controls per case will increase the power of the study and normally five controls per case is sufficient. But if cases are available and cost the same as controls, the most efficient size is typically to have the same number of cases and controls.

2.1.4.2. Binary Exposures Adjusted

To estimate sample size and power for detecting an association between a binary exposure X and disease (Y) association while controlling for confounders, PASS for logistic regression can be used. Now using the notation in PASS, power and sample size depend on α and β described in example 2.1; the percentage of nonexposed that are cases P0 and the percentage of exposed subjects that are cases P1; the proportion of exposure in the sample Pcnt N of X=1, which needs to be calculated based on P0 and P1 and the ratio of sizes between cases and controls: when equal cases and controls are assumed, (Pcnt N of X=1)=(1–2P0)/(2P1–2P0) × 100; and the R^2 ($0<=R^2<1$) achieved when X (considered as a continuous linear outcome) is regressed on W_1 W_2,…,W_K. A high value of R^2 indicates redundancy between X and {W_1 W_2,…W_K} in the model and therefore less power *(16)*. If only a single confounding covariate W_1 is involved, this R^2 is equal to the square of the Pearson correlation between X and W_1. For example, a R^2

Table 13.3
Two-proportion power analysis (based on approximation test) in PASS. Numeric results: Null hypothesis, P1 = P2; alternative hypothesis: P1<>P2

Power	N1	N2	Allocation ratio	P1	P2	Odds Ratio	α	β
0.80031	97	97	1.000	0.4000	0.6000	2.250	0.05000	0.19969

of 0.09 corresponds to the Pearson correlation of 0.3 between X and W_1 and a R^2 of 0.25 corresponds to the Pearson correlation of 0.5.

Example 2.2: If

- P0 = 40% of nonexposed group are cases and P1 = 60% of exposed group are cases (this corresponds to an OR of 2.25);
- A two-sided hypothesis test with $\alpha = 0.05$ will be used;
- Equal number of cases and controls are assumed; therefore, Pcnt N of X=1=50;
- $R^2 = 0.09$ or 0.25 between exposure and W_1 W_2, … W_K;
- N = 194, the total number of cases and controls.

The following output from PASS shows that the study will have 76% power when $R^2 = 0.09$ and 68% power when $R^2 = 0.25$ to detect an exposure OR of 2.25 (**Table 13.4**).

2.1.4.3. Continuous Exposures Unadjusted

If exposure is a continuous variable, then for unadjusted analyses a *t*-test is used. Using notations from PASS, power is now a function of α, β, N1, N2, the standard deviation for the exposure for controls S1, and for cases S2, the mean values for controls and cases under the alternative.

Example 2.3 If

- Controls have a mean value of a biomarker Mean1 = 150 and cases have a mean value of Mean2 = 170, with both cases and controls having the same standard deviation S1=S2=60;
- A two-sided hypothesis testing at $\alpha = 0.05$ will be used;
- Equal numbers of cases and controls will be used, that is, allocation ratio = 1.0 to achieve 80% power to detect a significant association between exposure and disease, N1=142 controls and N2=142 cases are needed (**Table 13.5**).

2.1.4.4. Continuous Exposures Adjusted

When the exposure variable is continuous and confounding variables need to be adjusted, the estimation of sample size and power also can be based on the logistic regression approach.

Table 13.4
Logistic regression power analysis in PASS

Power	Pcnt N	X=1	P0	P1	OR	R^2	α	β
0.76041	194	50.000	0.400	0.600	2.250	0.090	0.05000	0.23959
0.67638	194	50.000	0.400	0.600	2.250	0.250	0.05000	0.32362

Table 13.5
Two-group *t*-test power analysis in PASS. Numeric results for two-sample *t*-test: null hypothesis: Mean1 = Mean2. Alternative hypothesis: Mean1 <> Mean2

Power	N1	N2	Allocation ratio	α	β	Mean1	Mean2	S1	S2
0.80199	142	142	1.000	0.05000	0.19801	150.0	170.0	60.0	60.0

2.2. Matched Case-Control Studies

2.2.1. Why and Why Not to Match?

Another way to control for confounders is to use stratification at the design stage of a study. A form of such stratification occurs in case-control studies in which each case is matched with one or several controls on certain important confounding factors, such as age, sex, residence, etc. By matching on important confounders (as opposed to not matching), we can gain in statistical efficiency. For example, if we select general population controls at random in a lung cancer case-control study then the cases will be mostly "old" and the controls will be mostly "young." It will, therefore, be difficult to stratify on, and control for, age or in particular identify any factors other than age that influence lung cancer.

There are two types of matching: frequency matching and individual matching. Frequency matching (also called stratified matching) is matching the collective samples of cases and controls in terms of the collective distribution of the confounders of interest. It is often used when the matching variables are binary (such as sex) or have a small number of categories. Individual matching is to specifically match each case with one or more controls based on the matching criteria. It is often used when one of the matching variables are continuous (such as age) or have many levels of categories. Special statistical methods described below must be used to analyze individual matched case-control data.

Matching can be costly and time-consuming and furthermore matching on a weak risk factor (or a nonrisk factor) that is strongly correlated with the main exposure can dramatically reduce efficiency. Therefore, it is recommended only to match on variables that are easy to match, are not of intrinsic interest in themselves and are "soft" variables such as neighborhood or occupation that are difficult to adjust for in the analysis. Furthermore, it is recommended not to match on more than three different criteria because it is usually hard to find matches if too many criteria are used.

2.2.2. Statistical Inference on Exposure–Disease Association

A simplest analysis of matched data is with a 1:1 pair matching of cases with controls and a single binary exposure. The data can be summarized in **Table 13.6**,

Table 13.6
A 1:1 pair matching of cases with controls and a single binary exposure

No. of matched pairs

	Control	
Case	Exposed	Unexposed
Exposed	n_{11}	n_{10}
Unexposed	n_{01}	n_{00}

where the number in each cell represents the number of *matched pairs* in which the case and control are in the given exposure category. For example, if $n_{10} = 20$ there are 20 matched pairs for which the case was exposed and the matched control was not exposed. The statistical inference on exposure–disease association uses only the discordant pairs, that is, in which one but not both of the case and control are exposed. Conditional on the total number of discordant pairs, $n_{10} + n_{01}$, the number of discordant pairs where the case is exposed, n_{10} follows a binomial distribution, bin $(n_{10}+n_{01}, p)$ where the OR between exposure and disease, $\Psi=p/(1-p)$. Under the null hypothesis of no exposure–disease association, $p=1/2$. Thus, the test of the null hypothesis is based on testing the proportion of discordant pairs in which the case is exposed *(17)*. The OR for exposure–disease association can be estimated by

$$\hat{\psi} = \frac{n_{10}}{n_{01}}.$$

The estimate is known as the M-H estimate for matched pair data. The OR obtained from a matched case-control studies has the same interpretation as that calculated from a case-control study while adjusting for the matching variables. For example, if cases and controls are matched on age and gender, then the OR is the OR associated with the exposure after adjusting for age and gender. The confidence interval for Ψ is calculated by transforming the confidence limits for the binomial proportion p *(18)*. These methods for statistical inference for matched pairs can be generalized (using approaches developed by Mantel & Haenszel) to the setting where more than one controls are matched to each case.

If the exposure is continuous, then in 1:1 matched studies the outcome of interest is $d_j = X_{C,j} - X_{MC,j}$ where $X_{C,j}$ is the level of exposure in the case and $X_{MC,j}$ is the level of exposure in the matched control of the jth matched set. Hypothesis tests of no association between exposure and outcome can be performed using a paired t-test for $d_j = 0$. This can be extended to many

controls matched to one case by using analysis of variance models blocked by matching set *(19)*. Alternatively, and particularly if estimates of OR are important, associations with a continuous exposure variable in a matched case-control study can be tested using a conditional logistic regression described below. Multivariate conditional logistic regression also allows the matching variables to be effect modifiers and allows other confounding factors that have not been matched for to be controlled in the model.

2.2.3. Multivariate Analysis: Conditional Logistic Regression

In a one to one matched case-control study, the probability of being the case in the jth matched set given the covariate x can be modeled as

$$\text{logit}(p_j) = \alpha_j + \beta\chi.$$

where α_j, the intercept for the jth set, depends on the matching criteria and β is the coefficient for the effect of the exposure x. This formulation requires a large number of α_j's to be estimated, which can lead to seriously biased estimates of β *(20)*. However, the α_j's are nuisance parameters which can be eliminated using a conditional likelihood *(7, 21)*. Suppose in the jth matched set x_{1j} is the covariate for the case, say, "Jim" and x_{2j} is the covariate value for the control, say, "John", then as per the study design, conditional on exactly one of the two persons in the pair having the outcome, the probability that "Jim" is the observed case is as follows:

$$\frac{e^{\beta\chi_{1j}}}{e^{\beta\chi_{1j}} + e^{\beta\chi_{2j}}}.$$

This formulation easily extends to one case matched with more than one control or even multiple cases matched to multiple controls. As with logistic regression parameter estimates, variance and confidence limits under asymptotic normality for the conditional likelihood formulation are obtained using maximum likelihood estimation methods.

In a conditional logistic regression, the effect of the matching variable(s) on the risk of disease cannot be estimated, however, its interaction with an unmatched variable can be estimated by adding a product term between the matching variable and the exposure. For example, if the cases and controls are matched on one variable w_1, and to evaluate the possible effect modification by w_1, an interaction term in the conditional likelihood is as follows:

$$\frac{e^{\beta\chi_{1j}+\delta\chi_{1j}w_{1j}}}{e^{\beta\chi_{1j}+\delta\chi_{1j}w_{1j}} + e^{\beta\chi_{2j}+\delta\chi_{2j}w_{1j}}}.$$

Conditional logistic regression can be implemented using PROC LOGISTIC *(15)* and PROC PHREG *(22)* in SAS.

2.2.4. Sample Size and Power Estimation

2.2.4.1. Binary Exposures

Power and sample size are function of α, β, the number of matched pairs N, the probability of exposure for controls P0, the OR under the alternative ψ and the correlation of exposure between the case and matched control ϕ. Only positive correlations are considered. A value of zero indicates independence. The higher the correlation is, the less the power the study has. When no previous data are available for ϕ, Dupont *(23)* empirically suggests using a value of 0.2.

Example 2.4

Suppose in Example 2.1, the cases and controls are individually 1:1 matched on age and gender. If

- P0=40% of matched controls are exposed and $\Psi = 2.25$;
- Number of cases N = 97 and equal number of controls (controls per case M=1);
- Two-sided hypothesis test with $\alpha = 0.05$;
- $\phi = 0.2$.

Table 13.7 shows the matched case-control has 70% power to detect an OR of 2.25.

For continuous exposures, approaches for paired t-tests can be used.

2.3. Two-Stage Designs

In some epidemiological studies, a regular case-control study design is not feasible but a modified approach is necessary. For example, the disease status and exposure of primary interest in some epidemiological studies can be ascertained for a large group of subjects. However, information about important confounders can be collected only for a smaller sample because of either the cost or the difficulty in obtaining the information. A situation like this often arises in occupational studies. Suppose that employment records and basic medical records are available for

Table 13.7
Matched case-control power analysis

Power	Cases (N)	Controls/ case (M)	OR	Prob. exposed (P0)	Correlation (Phi)	α	β
0.69815	97	1	2.25	0.40000	0.20000	0.05000	0.30185

all employees so that their occupational exposure to a chemical agent is relatively easy to determine. However, the personal interviews needed to determine smoking and drinking history can only be conducted for a subsample of workers. A traditional case-control analysis would not be able to use all the exposure information because although exposure status is always available, smoking and drinking history is not available for every subject in the case-control study. A two-stage case-control design is thus applied instead in which the first stage the exposure and outcome information available for each case and control is used and in the second stage the confounder information collected only for a subset of cases and controls *(24,25)* is used. A unified analysis method that combines information from these two stages was developed by Cain and Breslow *(26)*. The idea is to use a standard logistic regression with the data collected at the second stage to estimate the exposure and disease association and then make simple adjustments to this estimate and its covariance based on data from stage one. Other statistical methods are also available for this setting *(27,28)*. Optimal sampling strategies in selecting sample size for two stages have been proposed by Reilly *(29)* and Hanley et al. *(30)*.

Because the cost of genotyping a large number of SNPs in a large case-control study may be prohibitive, another type of two-stage design has been adopted in which in the first stage, all the SNPs are typed on a subset of sample and only the SNPs that show an indication of association with disease are selected in the second stage to be typed on the remaining samples *(31)*. This type of two-stage sampling design is commonly used in gene association studies to reduce the cost of genotyping.

3. Design and Analysis of Cohort Studies

3.1. Time to Event Data

3.1.1. Introduction

The outcome of interest in the analysis of time to event data is the time until an event occurs. Time can be years, months, weeks, or days from the beginning of the follow-up of an individual until an event occurs. The event can be death, disease occurrence, relapse from remission, cure of cancer or any designated experience of interest that may happen to an individual. The exact time to event data is sometimes unobserved due to censoring. The most common type of censoring is called right censoring due to either subject loss of follow-up, subject withdrawal from the study, or the end of the study before the event occurs. However, time to event data can also be subject to left censoring (event is known to have occurred before an observed time) or interval censoring (event is known occurred within a given interval) *(32,33)*. An essential

assumption in most statistical approaches dealing with censored data that is often difficult to verify in practice is that censoring is independent of the outcome. Violation of this assumption leads to invalid results without proper adjustment. For example, in a prospective study on the natural history of HIV patients, if some patients leave the study when they become really sick hence their deaths are not observed, censoring is positively associated with death. Failure to adjust for noninformative censoring here can result in an underestimate of the death rate. Events also can be subject to left-truncation, which differs from left censoring in that left truncation removes the person entirely from the population. For example, one plans a prospective study to evaluate age-specific rate of colon cancer. A person who already had a colon polyp at the enrollment that was not diagnosed before is left censored (we know the polyp developed before the enrollment but do not know the exact date). However, people who died of colon cancer before the enrollment are left-truncated because they are removed from the population. Although the statistical approaches discussed here are designed for right censoring modifications exist for left censoring and left truncation.

Time to event data is characterized by survival functions. The survival function is defined as the probability of survival beyond certain time. The Kaplan–Meier approach gives a nonparametric estimate of survival function *(34)*, and the Greenwood formula estimates the variance of K-M estimates *(35)*. The Kaplan–Meier estimate has been shown to be less efficient than parametric estimates of the survival function *(36)*. However, Meier, et al. *(37)* recently demonstrated that this loss in efficiency is minimal. The parametric estimates can be biased and the Kaplan–Meier estimates therefore are generally preferred. Comparison of survival curves between two or more groups can be implemented using the log-rank test. The log-rank test gives equal weights on earlier and later separations between survival curves. Other tests such as Wilcoxin, Taron–Ware, etc., allow different weights to be put on earlier and later separations of survival curves. These methods are discussed and compared in Lawless *(38)*, Kalbfleisch and Prentice *(33)*, and Collett *(39)*.

Often times an individual can experience only one of several different types of events over follow-up. For example, when the outcome of interest is heart disease, if a person dies of cancer at age 63, then he or she is precluded from ever getting heart disease. This type of survival data is called time to event data with competing risks. A typical approach for analyzing such data is called the "cause-specific approach," in which a survival analysis is performed for each event type and the other competing event types are treated as censored. In the above-mentioned heart disease example, the person would be censored from the analysis on heart disease when he or she dies of cancer. However, this

approach requires the assumption that the occurrence of competing events is independent (or conditionally independent after considering all other covariates in the model) of the likelihood of the occurrence of the event of interest. This may often not be the case; in the above-mentioned example the person may have died of cancer in part because of poor health and for that reason would otherwise have higher than average risk for heart disease. Although several approaches have been developed for competing risks, implementation of these approaches has been hindered by nonidentifiability *(40)* and interpretability problems.

3.1.2. Grouped Cohort Data

A traditional method to analyze time to event data is the person year analysis in which the time to event is summarized in person-years *(41)*. Most of the essential concepts of person-year analysis can be introduced by considering a simple example of a two-dimensional table of rates. Suppose there is only one exposure variable with K levels, for example, history of smoking (none, light, medium, and heavy smoking) and confounding variables, for example, age and gender, are grouped into J strata. Let d_{jk} denote the number of death occurred at exposure level k and stratum j and assume $d_{jk} \sim Poisson(\lambda_{jk} n_{jk})$ where n_{jk} is the summation of time to event or follow-up time for all subjects at exposure level k and stratum j where k=1,…,**K** and j=1,…,J and λ_{jk} is the corresponding rate of death. A commonly used model for λ_{jk} is a multiplicative model (which is additive on the log scale):

$$\log \lambda_{jk} = \alpha_j + \beta_k$$

where $\alpha_j = \log \lambda_{j1}$ and β_k is the log of the death rate (or hazard) ratio between the k^{th} and first level of exposure while adjusting for confounders. This is called a log-linear model. The log-linear model can be implemented using PROC GENMOD in SAS. In a log-linear model, the exposure variable and the confounders have to be categorical. Continuous variables such as number of cigarettes per day or age in years cannot be used directly and have to be categorized. Furthermore, the rate for event is assumed to be independent of time, that is, the rate for event remains constant for the entire period that each individual is under observation. Constant rate for event is often not appropriate for clinical and public health applications as the risk for event may change with time on the treatment or age. The Cox proportional hazards model overcomes these limitations and therefore is more commonly used than the log-linear model.

3.1.3. Cox Proportional Hazards Model

Time to event data also can be characterized by hazard functions. The hazard function at time t, $\lambda(t)$, is defined as the instantaneous failure rate conditioning on surviving to time t. The Cox proportional hazards model *(42,43)* is defined based on a hazard function and allows evaluation of the effect of multiple covariates,

including continuous covariates on the risk of event. The most commonly used form of Cox model uses a multiplicative hazard function as follows:

$$\lambda(t) = \lambda_0(t)e^{\beta_1 x + \gamma_1 w_1 + \ldots + \gamma_k w_k}$$

where $\lambda_0(t)$ is the baseline hazard function (with all covariates set to 0) and $\exp(\beta_1)$ is interpreted as the ratio of instantaneous failure rate between x and x+1 holding W_1 W_2, …. W_K constant. The Cox model is semiparametric because the baseline hazard function is left completely unspecified; therefore, it can be of any shape, whereas the hazard ratio is assumed to be a parametric function of the β's. The Cox proportional hazards model allows covariates to be time-dependent. Time-dependent variables are the variables whose values are allowed to change with time. For example, the risk for AIDS among HIV+ patients depends on the current value of the CD4 count. Because CD4 count changes over time, it is important to include CD4 count as a time-dependent covariate in the model. The Cox model can be implemented using PHREG in SAS *(22)*. The time-dependent variable can be introduced using counting process notations *(44)*.

The key assumption in the Cox model is the proportionality of hazards for each covariate in the model. This assumption of proportionality of hazards can be assessed using graphical plots, a chi-square test on Schoenfeld residuals *(45)* or by the inclusion of time-dependent variables. An indication of violation of the proportionality assumption suggests invalidity of the application of the Cox PH model. However, the use of stratified Cox models or the inclusion of time-dependent covariates often resolves the issue of nonproportionality.

Time to event data also can be quantified on an age metric according to the age of an individual when an event occurs. In epidemiological cohort studies, age is often a strong determinant of disease. However, time on study instead of age is commonly used as the time scale. Thiebaut and Benichou *(46)* have shown via simulations that when the survival function is a function of age, that is, when age is the true time scale, using time on study as a time scale can induce bias in hazard ratio estimate for the covariate of interest even if age is adjusted for in the Cox model. They therefore strongly recommended not using time-on-study as the time scale for analyzing epidemiological cohort data. Because in certain situations, time-on-study as the time scale may be of interest such as short-term study or occupational studies, or when age itself is a prognostic factor of interest, we recommend using both time scales and comparing the results. When using age as the time scale, time to event data is left truncated. Most statistical software allows Cox analysis of left-truncated data by specifying age at enrollment and age at event or end of the study for each subject.

Often survival times are not observed more precisely than within an interval. For example in a study of HIV infection of babies of HIV-positive mothers where the babies are tested at 0, 3, 6, 9 … months, the exact time of HIV infection is not known but rather that infection occurred: before birth, or somewhere during postbirth intervals (e.g., 0–3 months, 3–6 months, 6–9 months, …). Survival data of this form are known as grouped or interval-censored data. Discrete analogues of the continuous proportional hazards model have been used to investigate the relationship between these survival times and a set of explanatory variables *(47,48)*. When the length of each interval is constant, grouped survival models can be reduced to a longitudinal binary response model with the complementary log-log link function or the logit link function *(49)*. Other authors have developed interval-censored models for settings where the intervals are not of constant length or position *(50)*, but these models are often difficult to apply in practice.

3.1.4. Accelerated Failure Time Model

Accelerated failure time models are another useful regression approach for time to event data *(38,43)*. For example, accelerated failure time models have been used in cardiovascular studies to model the time with successful aging *(51)*. The effect of a risk factor (i.e., covariate) in an accelerated failure time model is to change the scale of a baseline distribution of failure times: for a risk factor of interest X, the event time is $\varphi(\beta_\chi)T_0$, where $\varphi(0)T_0$ is the event time when the risk factor equals 0. Typically, an exponential parameterization is used for $\varphi(.)$, that is, $\varphi(\beta_\chi) = \exp(\beta_\chi)$. Thus, for $\beta < 0$, the risk factor shortens the duration of time; for $\beta > 0$, the risk factor increases the duration of time successful aging is maintained. The acceleration factor $\exp(\beta)$ estimates the impact on relative duration of successful aging from increasing in X by one unit. Although no proportional hazards assumption is required in an accelerated failure time model, a parametric distribution function is assumed for T_0. Common choices include the Weibull, Lognormal, and Gamma distributions. A monotone baseline hazards function is assumed in the Weibull model; an inverted U-shaped hazards function is assumed in the Lognormal model, and a U-shaped hazards function in T_0 is assumed in the Gamma model. An accelerated failure time model also has proportional hazards when the time to event follows a Weibull distribution and when there are no time-dependent covariates. The accelerated failure time model can be implemented using PROC LIFEREG in SAS.

3.1.5. Sample Size and Power

3.1.5.1. Unadjusted analyses

To estimate sample size and power to assess an exposure/treatment on time to event, log-rank tests can be used if a binary exposure is being evaluated. Power and sample size are a function of, using PASS notation, α and β; the portions of treatments

and controls S1 and S2, respectively, surviving over a given time interval (Note: the length of the interval does not matter as long as it is the same for both groups); the total number of subjects that start the study N; the portion of these subjects in the control group and the loss to follow up before the end of the study including censoring of outcome by the end of study.

Example 3.1

To plan a clinical trial, we expect

- S1 = 64% in the treatment group and S2 = 50% in the control group do not develop recurrence of tumor within two years. The program internally assumes that the survival function follows an exponential distribution. Based on this assumption, the difference in proportion of survival between the two groups corresponds to HR of 1.55 between control and treatment groups;
- 50% of the subjects are in treatment group (equal size of two groups);
- Assume 10% loss of follow-up by the end of the study;
- Two-sided hypothesis test with α=0.05.

To achieve 80% power, the following output shows that N = 433 total subjects are needed with 217 (N1) treated and 216 (N2) controls (**Table 13.8**). Note that loss to follow-up is conservatively assumed to all occur at the beginning of the study in this power estimation.

3.1.5.2. Adjusted Analyses

The Cox PH model analog of the log-rank test can be used if the exposure is continuous or there are other covariates that also need to be adjusted. Now the relationship between power and sample size is defined as a function of α, β; the overall event rate P, that is, the proportion of deaths occurred during the duration of the study (the proportion of noncensored subjects). This event rate should already take into account loss of follow-up; the standard deviation of X, SD; B = log (hazard ratio/unit of X) and the correlation between X and the other variables in the model in a linear regression model R^2 as described in Section 2.1.4.2.

Example 3.2

Table 13.8
Log rank survival power analysis–simple numeric results with proportion lost to follow up = 0.1000

Power	N	N1	N2	S1	S2	Hazard ratio	Two-sided α	Two-sided β
0.8011	433	217	216	0.6400	0.5000	1.5531	0.0500	0.1989

We plan a prospective study to evaluate adjusted effect of a continuous covariate (e.g., CD4 count) on survival of HIV-infected persons. Assume a Cox PH model with CD4 count as the primary covariate of interest while adjusting other confounding variables such as age, race, and HIV viral load.

- Expect the log hazard ratio associated with one unit increment in CD4 count, $B = 0.002$ (corresponds to HR of 1.22 for ΔCD4 = 100);
- Assume the standard deviation of CD4, S.D. = 200;
- R^2 = 0.18 between CD4 and the other variables W_1 W_2 … W_K in a linear regression model;
- The overall event rate (P) is *30%*;
- Two-sided hypothesis test with α = 0.05.

The following output (**Table 13.9**) shows the sample size of 267 (N) is needed to achieve 90% power.

As we discussed in **Subheading 3.1.4.**, accelerated failure time models are the same as Cox proportional hazards models under the assumption of a Weibull survival function. Therefore, the above PASS's procedure to compute power and sample size for the Cox model can be applied to the accelerated failure time model for power calculations.

3.1.6. Multivariate Failure Time Data

Multivariate failure time data arise when study subjects experience occurrences of multiple events: for example, tumor progression and death; or multiple occurrences of the same event: for example, tumor recurrences or multiple infections after surgery; or when there exists some natural or artificial clustering of subjects that induces dependence among the failure times of subjects within the same cluster, for example, time to occurrence of blindness in the left and right eyes or time to tumor appearance for rats living in the same litter.

There are several approaches to analyze multivariate survival data. Wei et al. *(52)* developed a marginal approach (often referred as the WLW method). The WLW method is most suitable for multiple events data, in which each type of event is allowed to

Table 13.9
Cox regression power analysis

	Sample Size	Reg. Coef.	S.D. of X1	Event rate	R^2 X1 vs Other X's	Two-sided	
Power	(N)	(B)	(SD)	(P)	(R2)	α	β
0.90005	267	0.002	200.000	0.300	0.180	0.05000	0.09995

have its own baseline hazard function and to have its own hazard ratio estimate. The WLW model is defined as the follows:

$$\lambda(t) = \lambda_{0i}(t)e^{\beta_{i\chi}+\gamma_{i1}w_1+\ldots+\gamma_{iK}w_K}$$

where $\lambda_{0i}(t)$ is the baseline hazard function for the ith type of event and $\exp(\beta_{i1})$ is the hazard ratio for event i between two subjects with covariate x differing in value by one while holding others constant. The WLW model assumes an independence working assumption when estimating the hazard ratio for covariates and uses robust variance estimates to conservatively adjust for any errors in this assumption. For example, in a cohort study of HIV-seropositive women on their natural history of human papillomavirus (HPV) infection and immune status *(53,54)*, the association of immune status measured by CD4 and HIV viral load on incident detections of HPV infections was evaluated. Incidence detections of 43 types of HPV (HPV6, 11, 12, 16,18, etc.) were analyzed using the WLW model to adjust the results for possible correlations between multiple types of incident HPV infections. The baseline risk of HPV infection and the association of immune status with HPV are allowed to be different across HPV types. When they are assumed to be common across the HPV types, the WLW model reduces to the LWA model *(55)*. The LWA model is also suitable for clustered data when subjects within the same cluster, such as twins or left and right eyes, are not to be distinguished from each other.

The WLW model also can be applied to recurrent events data to evaluate, for example, an HIV treatment effect on each recurrence of HPV infections (i.e., time from enrollment to first occurrence of HPV infection, time from enrollment to second occurrence of HPV infection, etc). Recurrent event data also can be analyzed using the intensity model *(56)*, the proportional rates/means model *(57–59)* or the gap time model on time in between each recurrence *(60)*. All of these analyses can be implemented using PROC PHREG in SAS.

3.2. Cohort Studies for Repeated Measures

3.2.1. Longitudinal Data Analysis

Sometimes, the outcome of a cohort study is not a binary survival event, but rather is an ongoing life state that changes over time (e.g., weight, blood pressure, depression state, etc.). To this end, a longitudinal (cohort) study is a study that individuals are measured repeatedly for this life state usually at well-defined time points. We use the above HPV example described in **Subheading 3.1.6.** to illustrate a longitudinal study. In this study, HPV infection status and CD4 and HIV viral load were repeatedly assessed at 6-month visits during the follow-up. On each woman in the study CD4, HIV, and HPV-status measures were taken at baseline and at 6, 12, 18 … months after baseline. All of these measured markers can change in either direction over time; for example,

active HPV infection is not permanent, but can remit and reoccur later. The repeatedly accessed outcome can be binary such as HPV infection or continuous such as CD4 count and HIV viral load. The repeated measures from the same subject (i.e., at baseline, and 6, 12, 18 months, …) tend to be correlated with each other by sharing common within person errors. Appropriate statistical methods to analyze such data must take into consideration the possible within person correlation.

There are three major approaches to do this. We illustrate these three approaches with a binary outcome, but they also can be applied to continuous outcomes as well as count data. The first approach is called a marginal model. We assume Y_{ij} is a binary outcome variable (0 or 1) and X_{ij} is the exposure of interest and W_{ij1}, W_{ij2}, …. W_{ijk} are the potential confounders for the jth longitudinal measure of the ith person. Let P_{ij} denote the probability of Y_{ij} being 1 the marginal model is then

$$\text{lo git}(P_{ij}) = \beta_0 + \beta_{1\chi_{ij}} + \gamma_1 w_{ij1} + \ldots + \gamma_K w_{ijK}.$$

To use the above HPV example to illustrate the model, if Y_{ij} is the HPV status (1 = ongoing HPV infection, 0 = no ongoing active HPV infection) for ith person at the jth visit and x_{ij} is her CD4 count at that visit, then β_1 is the log of odds ratio for HPV infection between two randomly selected women from the population whose CD4 counts value differ by 1 holding other variables constant. Parameters for marginal models are most conveniently estimated by the Generalized Estimating Equation (GEE) approached proposed by Liang and Zeger *(61)*. In GEE, the covariance of the parameter vector β is estimated based on a "robust" covariance by specifying a working correlation between repeated measures of Y. Commonly used choices of working correlations include independence, exchangeable, autoregressive or completely unspecified correlations. The independence assumption assumes no correlation between repeated measures and is more suitable when a small correlation is expected, whereas the exchangeable assumption assumes the same correlation between any pair of observations and is most suitable for sibling data. The autoregressive correlation assumes that the correlation between any pair of observations depends how far apart in time these observations are (the further they are apart the smaller the correlation is) and is often best suited for longitudinal data. However, even if the working correlation is not correctly specified, the variance estimates obtained from GEE is robust to this misspecification of the working correlation, meaning confidence interval for β and other statistical assessments of the model will be asymptotically conservative (the true type-1 error will be smaller than the nominal type-1 error). Choosing the working correlation to be close to the actual correlation,

however, increases the efficiency of the model. In practice, it is recommended to use several choices of working correlations and compare the results. GEE can be extended to multinomial data when the response variable is a nominal level variable with >2 levels *(62,63)*.

The second approach for longitudinal correlated data called a mixed effects model tries to more accurately model the time dependent processes and individual effects rather than rely on using robust variance for misspecification of the dependence structure. A simple random effects posits

$$\text{lo git}(P_{ij}) = \mu + \alpha_i + \delta_1 x_{ij} + \eta_1 w_{ij1} + \ldots + \eta_K w_{ijK}$$

which allows each subject to have his (or her) own starting point $\mu + \alpha_i$ and α_i is commonly assumed to follow a normal distribution, i.e., N (0, Ω). Again, we use the HPV example to illustrate the model: here $\mu + \alpha_i$ is person i's baseline risk of HPV infection and δ_1 is the log of odds ratio for HPV infection if a subject changes her CD4 count value by one unit with the assumption that increasing CD4 by unit has the same effect on the log-odds for all persons. The mixed effects model also can allow each subject have its own log-odds ratio associated with the disease. The mixed effects model is estimated using either the restricted maximum likelihood estimate or the maximum likelihood estimated where unlike GEE, it is important to specify all of the components correctly because estimation is not robust. However, because the covariance is nonrobust, the mixed effects approach is more efficient than marginal models.

When the outcome Y is a continuous variable such as the CD4 count, the marginal model is $Ey_{ij} = \beta_0 + \beta_1 x_{ij} + \gamma_1 w_{ij1} + \ldots + \gamma_K w_{ijK}$ and a simple mixed effects model is $Ey_{ij} = \mu + \alpha_i + \delta_1 x_{ij} + \eta_1 w_{ij1} + \ldots + \eta_K w_{ijK} + \varepsilon_{ij}$, where ε_{ij} is the within subject variation with $\varepsilon_{ij} \sim \mathbf{N}(0, \sigma^2)$. The marginal model can be analyzed using PROC GENMOD. For the mixed effects model, when Y is continuous, PROC MIXED in SAS can be used to implement such model. When Y is not continuous such as binary, PROC NLMIXED and PROC GLIMMIX in SAS can be used to implement this model. Note that the β in the marginal model is in general smaller in magnitude than the δ in the mixed effects model, that is, the population-averaged effect is in general smaller than the subject-specific effect. But β = δ when Y is continuous. A detail discussion on comparison of these two models can be found in refs. *64* and *65*.

The third approach for the analysis of longitudinal data is called the transition model. It models the current response as a function of past responses as well as exposure and confounding variables. The goal is to use previous responses to model the time dependence between responses explicitly so that a regular regression model (assuming independence between responses) can be

applied by simply treating previous responses as additional covariates in the regression model. The most commonly used transition model is first order or second order Markov models, which in addition to the covariates, the previous one or two responses Y_{ij-1} and Y_{ij-2} are used to predict the most recent response Y_{ij}. For example, if the HPV infection at the jth visit not only depends on this person's CD4 count at this visit, but also one's previous 2 measures of active HPV infections (which could for example indicate inactive latent HPV), the Markov model is then

$$\text{logit}(P_{ij}) = \beta_0 + \beta_{\chi_{ij}} + \gamma_1 w_{ij1} + \ldots + \gamma_K w_{ijK} + \alpha_1 y_{ij-1} + \alpha_2 y_{ij-2}.$$

When the Markov model is misspecified, the use of robust variance that we discussed for the GEE model will often give consistent estimate for the variance *(65)*.

Longitudinal data is most commonly collected prospectively by the follow up of a cohort, but it also can be identified retrospectively, by extracting multiple measurements taken over time on each person from historical records. However, a retrospectively collected longitudinal study may be inferior in terms of the quality of the data *(66)*. Statistical methods in the analysis of longitudinal data apply to both situations.

3.2.2. Sample Size and Power Approximation

Models for longitudinal data analysis are often too complex and cumbersome to derive an exact power estimate for the model. Simulation studies are, therefore, sometimes necessary to estimate power for such studies. Here we present a simple approach, which may be used as a first order approximation for power and sample size of a longitudinal study.

One approach to estimate a "ball park" sample size for longitudinal studies, is to first estimate the number of subjects needed to achieve 80% power if each subject only contributes one observation, say N. Then if each subject is contributing ~m repeated observations, the number of subjects needed will be less than N but will be greater than N/m because the repeated observations are correlated with each other. If the average (c.f. residual) correlation among all the repeated measures of a subject is ρ, then the number of subjects needed are $N(1+(m-1)\rho)/m$. The following table provides the ratio of number of subjects needed if each subject only contributes one observation to the number of subjects needed if each subject contributes m repeated observations to achieve the same power.

For example, if 200 subjects are needed with one measure per subject and ρ=0.3, with ~5 repeats per subject then 200/2.3 = ~87 subjects are needed for this study. **Table 13.10** also shows that when the correlation between repeated observations is small (0.3), having repeated observations will lead to considerable increases in power. But once the number of repeated observation

Table 13.10
Subjects needed with no repeats relative to with repeats

No. of repeats/ subject	Correlation between repeated Obs		
	0.3	0.5	0.8
3	1.8	1.5	1.2
5	2.3	1.7	1.2
10	2.7	1.8	1.2
20	3.0	1.9	1.2

reaches to a certain level, for example, 10, additional observations per subject lead to only minor increase in power, which may not be very cost-effective. When the (c.f. residual) correlation between repeated observations is high (above 0.5), power is not increased much by having repeated observations. This result can be applied to mixed effects model and marginal models for continuous data *(67)* and for binary data when an exchangeable correlation is suitable *(68)*.

4. Studies Nested within Cohort Studies

Ideally, when a cohort is available, it is best to use all individuals from the cohort for analysis. However, sometimes when expensive laboratory tests and/or limited samples are involved, this is not feasible and strategic subsets of cases and controls must be chosen for more detailed analysis. The two major approaches for this type of selections are nested case-control studies and case-cohort studies.

4.1. Nested Case-Control Studies

The nested case-control study is a case-control study generated within an established cohort to obtain estimates from a subsample of the cohort that are similar to estimates obtained from analysis of the entire cohort *(69,70)*. For each case the controls are chosen from those members of the cohort who are free of disease by the time of disease occurrence in the case. This type of sampling for controls is called incidence density sampling (or risk set sampling). Procedurally, incidence density sampling of controls

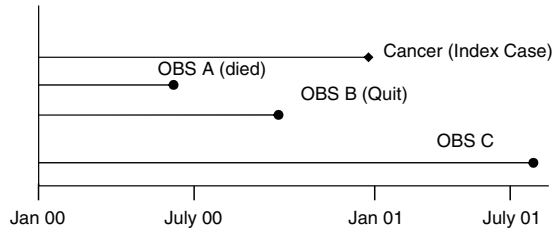


Fig. 13.1. Illustration of a nested case control study

is similar to traditional matching of controls: the case and the controls are matched on risk set defined by the case, that is, the controls have to be free of event at the time of disease occurrence in the case. We use a hypothetical example to illustrate incidence density based on time since enrollment.

Example 4.1

A cohort study of breast cancer among women who used and did not use estrogen therapy begins January 1, 2000. We wish to do a nested case-control study. A woman developed breast cancer on January 1, 2001, which we refer to as the index case. There are three other observations in the study (**Fig. 13.1**):

OBS A: Died on June 1, 2000

OBS B: Quit the study on November 1, 2000

OBS C: Alive and disease free on July 1, 2001

With incidence density sampling, Only OBS C is eligible to be matched to that index case. Note that if OBS C had developed breast cancer on June 1, 2001, she still could be matched to the index case because she had not developed breast cancer by January 1, 2001; if OBS A had developed breast cancer on June 1, 2000 (instead of dying) she still could not be matched to the index case because she had already developed breast cancer by January 1, 2001.

The risk set also can be defined based on the age of the case at the time of occurrence of the event, or a combination of age and time since enrollment. Here, we also use a hypothetical example to illustrate incidence density sampling based on age.

Example 4.2

We wish to do a nested case-control study for the association of Gene Q with heart disease in adults identified in a city registry (i.e., all people from that city are in the registry from birth). The index case is a person who developed heart disease at 74 years of age. Suppose there are three other observations in the study:

OBS A: Died at 73 years old

OBS B: Moved away from the city at 72 years old

OBS C: Alive without heart disease at 76 years old

With incidence density sampling, only OBS C is eligible to be matched to that index case. Note that if OBS C had developed heart disease at 75 years old, he would still be eligible to be matched to the index case because he did not have heart disease at 74 when the index case was diagnosed.

The advantage of the nested case-control study over the cohort study is that the nested case-control study uses only a sample of the cohort rather the entire cohort, therefore greatly reducing the labor and cost of data collection with a relatively minor loss in statistical efficiency *(71)*. The nested case-control study has been widely used in large cohort of patients from prospective studies and randomized clinical trials *(72–74)*. The number of controls selected per case may vary, but it is common in the nested case-control literature to find four or five controls per case. The nested case-control study is essentially an incidence density matched case-control study on time since enrollment or age and thus is analyzed using a conditional logistic regression. The sample size calculation for nested case-control study is also typically calculated the same way as a matched case-control study.

Counter-matching was introduced by Langholz and Clayton *(75)* as a method of sampling controls within each risk set for nested case-control studies. We illustrate this method using a simple case of one case matching with one control: if the index case is exposed, then it will be paired with an unexposed control; if the index case is unexposed, then it will be paired with an exposed control. The goal of counter-matching is to maximize the number of discordant case-control pairs. Counter-matching has been shown to increase the efficiency in estimating the effect of exposure by approximately 25% compared with classical random sampling *(76)*. However, one drawback of counter-matching is that it requires the exposure variable to be known for all the cohort members, whereas the nested case-control study with classical random sampling only requires the exposure variable to be known for all cases and the selected controls.

Counter-matching has also been shown to improve efficiency of nested case-control studies in testing interactions with exposure where the exposure variable is known for the entire cohort but possible confounders or effect modifiers are not practically assessed on all cohort members *(77)*. For example, in a case-control study of joint effect of atomic bomb radiation exposure and serum oestradiol levels (SEL) on breast cancer, radiation exposure is measured in the entire cohort but it is not practical to measure SEL for all cohort members. A counter-matching nested case-control study by counter-matching on radiation exposure will increase efficiency in estimating the interaction between radiation exposure and SEL compared with a nested case-control study with random sampling *(78)*. Counter-matching seems to be the most efficient design for detecting gene–

environment interactions when both genetic and environmental factors are rare *(79)*.

4.2. Case-Cohort Studies

A case-cohort study for failure time data consists of a subcohort of randomly selected individuals from the underlying cohort, and either all additional cases not in the subcohort *(80)* or a subset of the remaining cases *(81)*. Covariate information is collected on this sample rather than the entire cohort, which greatly reduces the cost and effort in collecting covariates yet with limited loss of efficiency *(82)* compared with a large-scale cohort study with a low outcome rate.

A case-cohort study therefore consists of two components: 1) the subcohort is a random sample of the entire cohort selected at baseline with probability ξ and without regard to the future outcome status, so the subcohort represents the person-time experience of the entire cohort. The subcohort will have some cases. But typically with a rare disease, most of the cases will not be in the subcohort; 2) all or a random subset of cases outside the subcohort. Data of a case-cohort study are treated as prospective, time-to-failure data, but special design features need to be taken into consideration in the Cox regression model. A case inside the subcohort is analyzed in the usual way as in the standard Cox model: if an individual has the event at time t, then this case and all available controls at time t define a risk set; the individual contributes person-time until the event occurs. A case outside the subcohort is not considered to be at risk until just before the event occurs and is not included in earlier risk sets *(80)*. Because the subcohort is a randomly sampled from the entire cohort, Barlow *(83)* proposed to weight the subcohort controls by the inverse of the sampling fraction ξ. For example, if $\xi = 5\%$, then each subcohort control has weight 20. Therefore, instead of a regular partial likelihood function, the parameter of interest, that is, the hazard ratio, is estimated by maximizing a so-called pseudo-likelihood function with different weighting schemes. The regression coefficient β is an unbiased estimator of the log of the relative risk of the full cohort. The score contributions from the pseudo-likelihood maximization are not independent because due to the oversampling of cases, martingale theory does not apply so either 1) a more complicated asymptotic variance *(82)* or 2) a robust variance estimator has to be used instead *(83)*. The use of case-cohort studies has thus been hampered by the computational burden of estimating the variance and lack of appropriate software. Therneau and Li *(84)* greatly simplified the computational issues by developing programs using standard software such as SAS (PHREG) and Splus (coxph). Case-cohort designs are thus now being increasingly used in cancer epidemiological studies *(85–87)*. In the Zeka et al. *(87)* study of occupational exposure to metalworking fluids (MWF) and the risk of upper aerodigestive tract cancers, the original cohort contains 31,100 workers

who worked at automobile plants in Michigan. Assessment of exposure history to MWF for all workers would have been costly and time-consuming. Furthermore, laryngeal, esophageal, and stomach cancers were examined separately in relation to exposure to MWF. Therefore, a case-cohort study design was applied and a subcohort of 3110 workers was selected. A positive association was found between MWF exposure and larynx cancer but not for esophageal and stomach cancers.

Methods for determining the power and sample size of case-cohort studies are limited due to the complexity in estimating variances for the relative risk estimates. Cai and Zeng *(88)* developed an approximate method to estimate sample size/power for a case-cohort study for a binary exposure variable and rare disease assumption with all cases selected. The power of their proposed log-rank type tests for case-cohort studies can be approximated by

$$Power = \Phi\left[Z_{\alpha/2} + \tilde{n}\theta\sqrt{\frac{p_1 p_2 p_D}{q + (1-q)p_D}} \right]$$

where α is the two-sided type 1 error of the hypothesis test, θ is the log hazard ratio for the two exposure groups, p_j (j=1,2) are the proportion of the two exposure groups in the population, $\tilde{n}$ is the total number of subjects in the subcohort, p_D is the proportion of the failures in the full cohort, and q is the sampling fraction of the subcohort. Kim et al. *(89)* recently showed that when the disease is rare, one can apply standard case-control approaches to estimate the sample size and power of case-cohort studies with a binary exposure variable and the results are very similar to Cai and Zeng's method. This implies that existing software for comparing two independent binomial proportions can be used to calculate sample size and power for case-cohort studies. Specially, the proportions of exposure are compared between the case group and the control group where the number of cases are $p_D\tilde{n}/q$ and the number of controls are $\tilde{n}(1-p_D)$ and the proportions for cases and controls are p_1 and $e^{\theta}p_1/(1+p_1(e^{\theta}-1))$, respectively.

4.4. Comparison between Nested Case-Control and Case-Cohort Designs

When the outcome is rare, both the nested case-control and case-cohort designs can provide economic estimates of relative risk parameters under the proportional hazards model, while requiring exposure and covariate information on only a small subset of the cohort *(90)*. Barlow et al. *(91)* showed that the case-cohort design may be more efficient in some applications than the nested case-control study provided there is not a high level of late entry and early censoring, although Lanholz and Thomas *(92,93)* reported a small to moderate advantage in statistical efficiency for the nested case-control study with substantial late entry and censoring. Below, we discuss further several practical issues for comparing these two study designs.

4.4.1. Multiple Outcomes

Compared with a nested case-control study, a case-cohort study provides a more efficient design to analyze multiple outcomes because the same cohort control group can be reused for each of the outcomes, whereas with a nested case-control study, different control groups may be needed for each outcome. The case-cohort design is particularly well suited for epidemiological studies in which various outcomes are evaluated with respect to a fixed set of risk factors. For example, in a case-cohort study of styrene exposure and ischemic heart disease, both acute myocardial infarction and chronic ischemic heart disease were examined as different outcomes and a significant association was found between styrene exposure and acute myocardial infarction *(94)*.

4.4.2. Time Scale

As we discussed in **Subheading 3.**, it may be beneficial for some epidemiological studies to use different time scales (time since enrollment vs. age). The case-cohort study can accommodate such changes because control selection is independent of case timing. In the nested case-control study, if the time scale changes from time since enrollment to age, the risk sets have to be reconstructed. The controls thus would have to be reselected based on the new risk sets. For example, if a case occurs at 3 years after enrollment and he or she is 75 years of age at that time, when the time scale is time since enrollment, the risk set for this case is all the subjects who survived at least 3 years after enrollment; when the time scale is age, the risk set is then all the subjects who is at least 75 years of age.

4.4.3. Subcohort

In the case-cohort study, the subcohort provides both a window on the entire cohort and the opportunity to assess quality control. The subcohort can be used to estimate the demographic and exposure distribution in the entire cohort. For example, the subcohort can be used to estimate allele frequencies for the entire cohort. The subcohort also can be used to estimate and model disease incidence and the comparison of disease incidence in this cohort with an external population. Due to incidence matching, controls in the nested case-control study may not provide a true picture of the underlying population particularly if the disease is not rare and there is significant amount of early right censoring.

4.4.4. Computational Burden

The computation for nested case-control study is straightforward as the conditional logistic regression is performed with matching on time. Robust standard errors are not needed for nested case-control studies. Likelihood ratio and score tests for parameters can be used. But computation for case-cohort study is challenging as it was based on a pseudo-partial likelihood and robust variance has to be used. However, with the simplification of Therneau and Li *(84)*, the standard software can be easily applied to estimate the relative risk parameter and its robust variance. The likelihood

ratio test cannot be used in a case-cohort study, the Wald test has to be used instead.

4.4.5. Counter-Matching

The nested case-control study can take advantage of counter-matching, whereas the case-cohort study cannot.

In summary, the choice of choosing one design over the other should depend on the factors that are specific to a particular situation.

References

1. Benjamin, Y., and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300.
2. Westfall, P.H., and Young, S.S. (1993) Resampling-based Multiple Testing, New York: John Wiley & Sons, Inc.
3. Hoh, J., Wille, A., Zee, R., et al. (2000) Selecting SNPs in two-stage analysis of disease association data: a model-free approach. *Am. J. Hum. Genet.* **64**, l413–7.
4. Hintze, J.L. (2001) PASS: Power and Sample Size Software. East Kaysville UT.
5. Rothman, K.J. (1986) Modern Epidemiology, Boston/Toronto: Little, Brown and Company.
6. Armitage, P., and Berry, G. (1990) Statistical Methods in Medical Research, London: Cambridge University Press,.
7. Breslow, N.E., and Day, N.E. (1980) Statistical Methods in Cancer Research, Volume I: The Analysis of Case-Control Studies, IARC Scientific Publications, No. 32, Lyon, France: International Agency for Research on Cancer.
8. Armitage, P. (1955) Test for linear trends in proportions and frequencies. *Biometrics* **11**, 375–386.
9. Mantel, N. (1963). Chi-square tests with one degree of freedom: extensions of the Mantel-Haenszel procedure. *J. Am. Stat. Assoc.* **58**, 690–700.
10. Mantel, N., and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl. Cancer Inst.* **22**, 719–748.
11. Robins, J.M., Breslow, N.E., and Greenland, S. (1986) Estimators of the Mantel-Haenszel variance consistent in both sparse data and large-strata limiting models. *Biometrics* **42**, 311–323.
12. Agresti, A. (2002) Categorical Data Analysis, 2nd edition, New York: John Wiley & Sons, Inc.
13. Hosmer, D.W, Jr., and Lemeshow, S. (2000) Applied Logistic Regression, 2nd edition, New York: John Wiley & Sons, Inc.
14. Allison, P.D. (1999) Logistic Regression Using the SAS System: Theory and Application, Cary, NC: SAS Institute.
15. SAS Institute. (1995) Logistic Regression Examples Using the SAS System, Cary, NC: SAS Institute Inc.
16. Hsieh, F.Y., Block, D.A., and Larsen, M.D. (1998) A Simple Method of Sample Size Calculation for Linear and Logistic Regression. *Stat. Med.* **17**, 1623–1634.
17. McNemar, Q. (1947) Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* **12**, 153 – 157.
18. Liddell, F.D.K. (1983) Simplified exact analysis of case-referent studies: matched pairs; dichotomous exposure. *J. Epidemiol. Community Health* **37**, 82–84.
19. Ury, H.K. (1975) Efficiency of case-control studies with multiple controls per case: continuous or dichotomous data. *Biometrics* **31**, 643–649.
20. Cox, D.R., and Hinkley, D.V. (1974) Theoretical Statistics, London, UK: Chapman & Hall.
21. Stokes, M.E., Davis, C.S., and Koch, G.G. (2000) Categorical Data Analysis Using the SAS System, 2nd edition. Cary, NC: SAS Institute.
22. Allison, P.D. (1995) Survival Analysis Using the SAS System: A Practical Guide, Cary, NC: SAS Institute.
23. Dupont, W. (1988) Power calculations for matched case-control studies. *Biometrics* **44**, 1157–1168.
24. Walker, A.M. (1982) Anamorphic analysis: sampling and estimation for covariate effects when both exposure and disease are known. *Biometrics* **38**, 1025–32.
25. White, J.E. (1982) A two-stage design for the study of the relationship between a rare

exposure and a rare disease. *Am. J. Epidemiol.* **115**, 119–28.
26. Cain, K.C., and Breslow, N.E. (1988) Logistic regression analysis and efficient design for two-stage studies. *Am. J. Epidemiol.* **128**, 1198–206.
27. Scott, A.H., and Wild, C.J. (1997) Fitting regression models to case-control data by maximum likelihood. *Biometrika* **84**, 57–71.
28. Chatterjee, N., Chen, Y.H., and Breslow, N.E. (2003) A pseudoscore estimator for regression problems with two-stage sampling. *J. Am. Stat. Assoc.* **98**, 158–68.
29. Reilly, M. (1996) Optimal sampling strategies for two-stage studies. *Am. J. Epidemiol.* **143**, 92–100.
30. Hanley, J.A., Csizmadi, I., and Collet, J.-P. (2005) Two-stage case-control studies: precision of parameter estimates and considerations in selecting sample size. *Am. J. Epidemiol.* **162**, 1225–1234.
31. Thomas, D., Xie, R., and Mulugeta G. (2004) Two-stage sampling designs for gene association studies. *Genet. Epidemiol.* **27**, 401–414.
32. Maddala, G.S. (1983) Limited-Dependent and Qualitative Variables in Econometrics, New York: Cambridge University Press.
33. Kalbfleisch, J.D., and Prentice, R.L. (1980) The Statistical Analysis of Failure Time Data, New York: John Wiley & Sons, Inc.
34. Kaplan, E.L., and Meier, P. (1958) Nonparametric estimation form incomplete observations. *J. Am. Stat. Assoc.* **53**, 457–481.
35. Greenwood, M. (1926) The errors of sampling of the survivorship tables, in Reports on Public Health and Statistical Subjects, no. 33. London: HMSO, Appendix I.
36. Miller, R.G., Jr. (1983) What Price Kaplan-Meier? *Biometrics* **39**, 1077–1081.
37. Meier, P., Karrison, T., Chappell, R., and Xie, H. (2004) The Price of Kaplan-Meier. *J. Am. Stat. Assoc.* **99**, 890–896.
38. Lawless, J.F. (1982) Statistical Methods and Methods for Lifetime Data, New York: John Wiley & Sons, Inc.
39. Collett, D. (1994) Modeling Survival Data in Medical Research, p. 23, London, UK: Chapman & Hall.
40. Tsiatis, A.A. (1975) Nonidentifiability aspect of the problem of competing risks. *Proc. N Y Acad. Sci.* **72**(1), 20–22.
41. Breslow, N.E., and Day, N.E. (1987) Statistical Methods in Cancer Research, Volume II: The Design and Analysis of Cohort Studies, IARC Scientific Publications, No. 82, Lyon, France: International Agency for Research on Cancer.
42. Cox, D.R. (1972) Regression models and life tables. *J. R. Stat. Soc. Ser. B* **20**, 187–220.
43. Cox, D.R., and Oakes, D. (1984) Analysis of Survival Data, London, UK: Chapman & Hall.
44. Andersen, P.K., Borgan, Ø., Gill, R.D., and Keiding, N. (1992) Statistical Models Based on Counting Processes, New York: Springer-Verlag.
45. Schoenfeld, D. (1982) Partial residuals for the proportional hazards regression model. *Biometrika* **69**, 239–241.
46. Thiebaut, A.C.M., and Benichou, J. (2004) Choice of time-scale in Cox's model analysis of epidemiologic cohort data: a simulation study. *Stat. Med.* **23**, 3803–3820.
47. Prentice, P.L., and Gloeckler, L.A. (1978) Regression analysis of grouped survival data with applications to breast cancer data. *Biometrics* **34**, 57–67.
48. Allison, P.D. (1982) Discrete-time methods for the analysis of event histories. In: Sociological Methods and Research, 15 ed. S . Leinhardt, San Francisco, CA: Jossey-Bass, 61–98.
49. D'Agostino, R.B., Lee, M.-L., Belanger, A.J., Cupples, L., Anderson, K., and Kannel, W.B. (1990) Relation of pooled logistic regression to time dependent Cox regression analysis: the Framingham Heart Study. *Stat. Med.* **9**, 1501–1515.
50. Sun, J. (2006) The Statistical Analysis of Interval-censored Failure Time Data, NY: Springer.
51. Newman, A.B., Arnold, A.M., Naydeck, B.L., et al. (2003) Successful aging: effect of subclinical cardiovascular disease. *Arch. Intern. Med.* **163**, 2315–2322.
52. Wei, L.J., Lin, D.Y., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distribution. *J. Am. Stat. Assoc.* **84**, 1065–1073.
53. Strickler, H.D., Palefsky, J.M., Shah, K.V., Anastos, K., Klein, R.S., Minkoff, H., Duerr, A., Massad, L.S., Celentano, D.D., Hall, C., Fazzari, M., Cu-Uvin, S., Bacon, M., Schuman, P, Levine, A.M., Durante, A.J., Gange, S., Melnick, S., Burk, R.D. (2003). Human papillomavirus 16 and immune status in human immunodeficiency virus-seropositive women. *J. Natl. Cancer Inst.* **95**, 1062–71.
54. Strickler, H.D., Burk, R.D., Fazzari, M., Anastos, K., Minkoff, H., Massad, L.S., Hall, C., Bacon, M., Levine, A.M., Watts,

H., Silverberg, M.J., Xue, X., Schlecht, N., Melnick, S., Palefsky, J.M. (2005). HPV Natural History and Possible HPV Reactivation in HIV-Positive Women. *J. Natl. Cancer Inst.* **97**, 577–86.

55. Lee, E., Wei, L., and Amato, D. (1992) Cox-Type Regression Analysis for Large Numbers of Small Groups of Correlated Failure Time Observations, Netherlands: Kluwer Academic Publishers, 237–247.
56. Andersen, P.K., and Gill, R.D. (1982). Cox's regression model counting process: a large sample study. *Ann. Stat.* **10**, 1100–1120.
57. Lin, D., Wei, L., Yang, I., and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *J.R. Stat. Soc. B* **62**, 711–730.
58. Lawless, J., and Nadeau, C. (1995) Some simple robust methods for the analysis of recurrent events. *Technometrics* **37**, 158–168.
59. Pepe, M., and Cai, J. (1993) Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *J. Am. Stat. Assoc.* **88**, 881–820.
60. Prentice, R.L., Williams, B.J., and Peterson, A.V. (1981). On the regression analysis of multivariate failure time data. *Biometrika* **68**, 373 – 379.
61. Liang, K.Y., and Zeger, S.L. (1986) Longitudinal data analysis using generalized linear models *Biometrika* **73**, 13–22.
62. Lipsitz, S.H., Kim, K., and Zhao, L. (1994) Analysis of repeated categorical data using generalized estimating equations. *Stat. Med.* **13**, 1149–1163.
63. Miller, M.E., Davis, C.S., and Landis, J.R. (1993) The analysis of longitudinal polytomous data: generalized estimating equations and connections with weighted least squares. *Biometrics* **49**, 1033–1044.
64. Zeger, S.L., Liang, K.-Y., and Albert, P.S. (1988) Models for longitudinal data: a generalized estimation equation approach. *Biometrics* **44**, 1049–1060.
65. Diggle, P.J., Liang, K.Y., and Zeger, S.L. (1994) Analysis of Longitudinal Data, Oxford: Clarendon Press.
66. Goldfarb, N. (1960) An Introduction to Longitudinal Statistical Analysis-the Method of Repeated Observations from a Fixed Sample, Glencoe, IL: Free Press.
67. Hoover, D.R. (2002) Power for t-test comparisons of unbalanced cluster exposure studies *J Urban Health* **79**(2), 278–94.
68. Pan, W. (2001). Sample size and power calculations with correlated binary data. *Controlled Clin. Trials* **22**, 211–227.
69. Kupper, L.L., McMichael, A.J., and Spirtas, R. (1975) A hybrid epidemiologic study design useful in estimating relative risk. *J. Am. Stat. Assoc.* **351**, 524–528.
70. Breslow, N.E., Lubin, J.H., Marek, P., and Langholz, B. (1983) Multiplicative models and cohort analysis. *J. Am. Stat. Assoc.* **78**, 1–12.
71. Ernster, V.L. (1994) Nested case-control studies. *Prev. Med.* **23**, 587–590.
72. Essebag, V., Genest J., Suissa S., and Pilote L. (2003). The nested case-control study in cardiology. *Am. Heart J.* **146**, 581–590.
73. Sidney, S., Friedman, G.D., and Hiatt R.A. (1986). Serum cholesterol and large bowel cancer. *Am. J. Epidemiol.* **124**, 33–38.
74. Krieger, N., Wolff, M.S., Hiatt, R.A., Rivera, M., Vogelman, J., and Orentreich, N. (1994) Breast cancer and serum organochlorines. *J. Natl. Cancer Inst.* **86**, 589–599.
75. Langholz, B., and Clayton, D. (1994). Sampling strategies in nested case-control studies. Environ. *Health Perspect.* **102**(Suppl 8), 46–51.
76. Steenland, K., Deddens, J.A. (1997) Increased precision using counter-matching in nested case-control studies. *Epidemiology* **8**, 238–42.
77. Langholz, B. (2005) Counter-matching. Encyclopedia of Biostatistics. 2nd edition. Vol.2. ed. P. Armitage and T. Colton, Chichester, UK: John Wiley & Sons, Ltd., 1248–1254.
78. Cologne, J.B., Sharp, G.B., Neriishi, K., Verkasalo, P.K., Land, C.E. and Nakachi, K. (2004). Improving the efficiency of nested case-control studies of interaction by selecting controls using counter matching on exposure. *Int. J. Epidemiol.* **33**, 485–492.
79. Andrieu, N., Goldstein, A.M., Thomas, D.C., and Langholz, B. (2001) Counter-matching in studies of gene-environment interaction: efficiency and feasibility. *Am. J. Epidemiol.* **153**, 265–274.
80. Prentice, R. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* **73**, 1–11.
81. Chen, K. (2001). Generalized case-cohort sampling. *J. R. Stat. Soc. B* **63**, 791–809.
82. Self, S.G., and Prentice, R. (1988). Asymptotic distribution theory and efficiency results for case-cohort studies. *Ann. Stat.* **16**, 64–81.
83. Barlow, W.E. (1994) Robust variance estimation for the case-cohort design. *Biometrics* **50**, 1064–1072.

84. Therneau, T.M., and Li, H. (1999) Computing the Cox model for case-cohort designs. *Lifetime Data Anal.* **5**, 99–112.
85. Ramadhani, M.K., Elias, S.G., van Noord, P.A.H., et al. (2005) Innate left handedness and risk of breast cancer: case-cohort study. *BMJ* **331**, 882–883.
86. Savitz, D.A., Cai, J, van Wijngaarden, E., et al. (2000) Case-cohort analysis of brain cancer and leukemia in electric utility workers using a refined magnetic field job-exposure matrix. *Am. J. Ind. Med.* **38**, 417–425.
87. Zeka, A,, Eisen, E.A., Kriebel, D, et al. (2004). Risk of upper aerodigestive tract cancers in a case-cohort study of autoworkers exposed to metalworking fluids. *Occup. Environ. Med.* **61**, 426–431.
88. Cai, J., and Zeng, D. (2004) Sample size/power calculation of case-cohort studies. *Biometrics* **60**, 1015–1024.
89. Kim, M.Y., Xue, X., and Du. Y. (2006) Approaches for calculating power for case-cohort studies. *Biometrics* **62**, 929–933.
90. Wacholder, S. (1991) Practical considerations in choosing between the case-cohort and nested case-control designs. *Epidemiology* **2**, 155–158.
91. Barlow, W.E., Ichikawa L., Rosner, D., and Izumi S. (1999) Analysis of case-cohort designs. *J. Clin. Epidemiol.* **52**, 1165–1172.
92. Langholz, B., and Thomas, D.C. (1990) Nested case-control and case-cohort methods of sampling from a cohort: a critical comparison. *Am. J. Epidemiol.* **131**, 169–176.
93. Langholz, B., and Thomas, D.C. (1991) Efficiency of cohort sampling designs: some surprising results. *Biometrics* **47**, 1563–1571.
94. Matanoski, G.M., and Tao, X. (2003) Styrene exposure and ischemic heart disease: a case-cohort study. *Am. J. Epidemiol.* **158**, 988–995.

Chapter 14

Methods in Cancer Epigenetics and Epidemiology

Deepak Kumar and Mukesh Verma

Summary

Cancer cells are characterized by epigenetic dysregulation, including global genome hypomethylation, regional hypo- and hypermethylation, histone modifications, and disturbed genomic imprinting. These alterations can be used as markers in cancer epidemiology to assess risk as well as follow-up of the disease. Some of the methylation markers can be assayed on microarray chips on which a very small amount of sample is needed, and the assays are robust and the markers have high specificity and sensitivity. Implication of epigenetic markers in cancer detection, diagnosis, prognosis, control, and cancer epidemiology is discussed.

Key words: Acetylation, biomarker, chromatin, epigenetics, histone, histone deacetylase (HDAC), histone acetyl transferase (HAT), methylation, methyl transferase, risk assessment, screening.

1. Introduction

Cancer is a genetic and epigenetic disease. After completion of the human genome project, there is a need to complete the human epigenome project *(1,2)*. Epigenetic processes are essential for the packaging and use of the genome, and they are fundamental to normal human development. Any abnormal process may contribute to cancer development *(3,4)*. The components of epigenome are DNA methylation patterns, histone modifications, chromatin modeling, and small RNA profiling. Although epigenetic changes are heritable in somatic cells, they are potentially reversible by drug treatments, and this has significant implications for major diseases, including cancer prevention and treatment. Due to alteration in the cell environment as a result of exposure to radiation, diet, or hormones, the components of the epigenome modify. These modifications lead to alterations in gene expression.

Mukesh Verma (ed.), *Methods in Molecular Biology, Cancer Epidemiology, vol. 471*
© 2009 Humana Press, a part of Springer Science + Business Media, Totowa, NJ
Book doi: 10.1007/978-1-59745-416-2

For example, hypermethylation of tumor suppressor genes results in their inactivation, and hypomethylation of oncogenes results in their activation (*5*). These modifications have great clinical diagnostic value and potential to be used as markers for screening populations to identify individuals with high risk of developing cancer. Here, we discuss in detail the procedures involved in detecting different epigenetic marks. Specific examples are included appropriately for clear understanding of epigenetic-mediated gene regulation during cancer development.

2. Materials and Methods in DNA Methylation

2.1. Materials for DNA Methylation

2.1.1. Cell Lysis and DNA Extraction

1. Lysis buffer containg proteinase K (Sigma-Aldrich, St. Louis, MO).
2. Phenol, chloroform, and alcohol for DNA extraction (Sigma-Aldrich).
3. DNA purification kit (QIAGEN, Valencia, CA).

2.1.2. Bisulfite Treatment

1. Sodium bisulfite (Sigma-Aldrich).
2. DNA purification kit (QIAGEN).
3. CpGenome DNA modification kit (Intergren, New York, NY).

2.1.3. Amplification of Methylated and Unmethylated DNA

1. Oligonucleotides specific for methylated and unmethylated regions (New England Biolabs, Ispwich, MA).
2. Polymerase Chain Reaction (PCR) machine (PerkinElmer Life and Analytical Sciences, Boston, MA).

2.1.4. Gel Electrophoresis

1. Restriction enzymes, ligase, ligation buffer, agarose, electrophoresis buffer (Sigma-Aldrich).
2. PerkinElmer sequence detection system.

2.1.5. Samples

1. Approximately 0.5 g of tissue is enough for methylation analysis. Some investigators have done methylation analysis in biofluids (e.g., blood, urine, saliva).

2.2. Methods for DNA Methylation

2.2.1. DNA Isolation

1. DNA is isolated by the standard method of proteinase K digestion and phenol chloroform extraction (or using any commercial kit) *(6,7)*.
2. Blood has emerged as the most common sample used in epidemiologic studies, because it is collected with relatively less invasive techniques, and most of the clinicians prefer diagnosis and prognosis based on blood sample analyses.

3. Exfoliated cells from other fluids (e.g., cerebral spinal fluid, seminal plasma, prostatic fluid, nipple aspirate, urine, saliva) also have been used for clinical studies.

2.2.2. Bisulfite Modification and Detection of Methylation Status of Multiple Loci

1. Sodium bisulfite modification is performed using a CpGenome DNA modification kit (Intergren).
2. DNA in 18-µl volume is denatured at 100°C for 10 min, and then it is centrifuged briefly, and chilled on ice.
3. NaOH is added to a final concentration of 0.3 M in 20-µl volume, and the sample is incubated at 42°C for 20 min.
4. Sodium bisulfite solution is prepared on the day of use by adding 1.9 g of sodium metabisulfite (Sigma-Aldrich) to 3.2 ml of 0.44 M NaOH and heating at 50°C to dissolve the bisulfite.
5. After addition of hydroquinonine (0.5 ml of a 1 M solution), 120 µl of this solution is mixed with each DNA sample.
6. The reaction is allowed to proceed at 50°C for 16 h in the dark.
7. Sodium bisulfite treatment of DNA converts unmethylated cytosines to uracil, whereas methylated cytosine residues remain unaffected.
8. This treatment creates methylation-dependent sequence differences in the DNA.
9. The methylation status of different genes is evaluated by combined restriction analysis and MethyLight technologies *(6,8)*.
10. Primers are selected to cover the CpG island. Sodium bisulfite-treated genomic DNA is amplified by fluorescence-based, real-time quantitative PCR.
11. DNA is amplified using locus-specific PCR primers flanking an oligonucleotide probe with a 5′-fluoresecent reporter dye and a 3′-quencher dye.
12. Because Taq has 5′ to 3′ nuclease activity, it cleaves the probe and releases the reporter and fluorescence of the released reporter can be detected by PerkinElmer sequence detection system.
13. After a threshold of amplified DNA is achieved, the product generates a fluorescent signal that is proportional to the amount of PCR product generated. Routinely, serial dilutions of a control sample are included on each plate to generate a standard curve. Plate-to-plate consistency is achieved by including several reference samples on each plate.
14. The amplification step is performed in a 96-well plate with a final reaction mixture of 25 µl consisting of 600 nM each primer ; 200 nM probe ; 200 µM each dATP, dCTP, and dGTP ; 400 µM dUTP ; 3.5 mM $MgCl_2$; 1× TaqMan buffer A and bisulfite DNA.

15. Amplification conditions are 50°C for 2 min, 95°C for 10 min, followed by 42 cycles at 95°C for 15 s and 60°C for 1 min. For controls, one unrelated fully methylated gene, one fully unmethylated gene, and internal reference set to check input DNA also are included.
16. The methylated and unmethylated primers and the probe were designed to overlap about five potential CpG dinucleotide sites.
17. For internal controls, parallel reactions are run with primers specific for the bisulfite-modified methylated or unmethylated sequences of genes under study and with MYOD1 or ACTB reference primers.

2.2.3. Characteristics and Advantages of MethyLight Technology for Detecting Methylated Genes

There are several techniques available for detecting methylated DNA sequences, such as COmbined Bisulfite Restriction Analysis (COBRA) and MethyLight technology. MethyLight technology has few specific features, described below, that make it an easy and sensitive technology *(7)*.

1. Sequence discrimination at the PCR amplification level occurs by designing the primers and probes, or just primers, to overlap potential sites of DNA methylation.
2. The MethyLight technology can be modified to avoid sequence discrimination at the PCR amplification level; if neither the primers nor the probe overlap any CpG dinucleotides, then the reaction represents unbiased amplification and can serve as a control to determine the input DNA.
3. When the probe is designed just to cover CpG islands, sequence discrimination occurs completely at the hybridization step.
4. There is no amplification bias.
5. No restriction digestion step is needed.
6. MethyLight technique is linearly quantitative over 4 orders of magnitude, this is a far greater range of quantitative accuracy than reported for any of the other DNA methylation analysis method to date.
7. MethyLight technique has been used in colon cancer epidemiology studies (hundreds to thousands of samples) to identify high-risk population.
8. This technique is compatible with paraffin-embeded tissue samples.
9. Unlike other techniques, assays in MethyLight are completed at the PCR level itself; thus, gel-based separation is not needed and it reduces labor and chances of sample contamination.

2.2.4. Bisulfite Sequencing

1. DNA is bisulfite converted and bisulfite sequencing and COBRA primers, designed specifically to exclude CpG dinucleotides, are used in PCR for 35 cycles.
2. For bisulfite sequencing, a cleanup step is used to purify the product (QiaPrep spin kit, QIAGEN).
3. The TOPO TA kit is used to clone purified products (Invitrogen, Carlsbad, CA). Sequencing from plasmid pCR2.1 is performed using the M13 forward primer with Big Dye Sequencing mix (Applied Biosystems, Foster City, CA) according to the instructions of the manufacturer.
4. Ten clones of each product are aligned and the percentage of CpG methylation was calculated for each site.
5. For COBRA, the product is digested by the enzyme indicated (all enzymes from New England Biolabs).

DNA methylation occurs mostly in the promoter region at 5-methylcytosine sites. CpG/CpG dyads are the principal sites of methylation. These dyads are composed of dinucleotide CpG on one strand and the complementary CpG nucleotide on the opposing DNA strand. Methylation pattern is maintained from parent to daughter starnds of DNA. Methylation occurs either de novo or methylation patterns are maintained. Specific methylating enzymes are needed for de novo methylation and methylation maintenance. Maintenance methyltransferase uses hemimethylated CpG/CpG dyads, when only one strand is methylated, and operated during DNA replication phase *(9)*. In contrast, de novo methyltransferase does not exhibit a preference for hemimethylated sites. It is still not clear how these two methyltransferases are regulated *(10)*.

Laird's group has demonstrated that de novo methylation can occur at a CpG site on the daughter starnd only and the parent strand CpG remains unmethylated *(3,4,9)*. They used hair-pin-bisulfite PCR on leukocyte samples. During de novo methylation only one strand got methylated, although it was immaterial which strand was considered, daughter strand or parent strand. Two genes, *FMR1* and *LINE-1*, were tested by this approach *(11)*.

Although methylation profiling is an efficient method, there exists an area that needs improvement in determining methylation pattern. Failure of bisulfite conversion is the main drawback of methylation detection methodologies. Sometimes two hemimethylated dyads are present in opposite polarity, but the frequency of such dyads is slow (e.g., 24 of 110 for *FMR1*). Careful experimentation has demonstrated that these dyads are not artifacts but that they occur due to altered gene regulation. Based on such studies, protocols have been established to follow the rate of methylation in real time. In a population-model–based study, where subjects were followed for up to 5 years, the general methylation profile of participants was similar at year 1 and year 5, and changes were observed only at different alleles at a single time point *(12)*.

3. Histone Modifications

3.1. Materials for Histone Modifications

1. Polyclonal Sera against Synthetic Peptides
2. Peptides (sequence depends on the properties of histone being studied).
3. IgG purification kit (Imgenex, San Diego, CA).
4. Phosphate-buffered saline (PBS), 1 M Tris-HCl, pH 8.8, Tween 20 (Pierce Chemical, Cramlington, Northumberland, UK).
5. NE buffer (20 mM HEPES, pH 7, 10 mM KCl, 1 mM $MgCl_2$, 0.5 mM dithiothreitol (DTT), 0.1% Triton, and 200 mM NaCl).
6. SDS-PAGE materials (GE Healthcare, Piscataway, NJ).

3.2. Methods for Histone Modifications

1. Chromatin immunoprecipitation assay is carried by lysing the cells and precipitating with specific antibodies.

3.2.1. Immunoprecipitation Assay

2. Cells are cross-linked by incubating them in 1% (v/v) formaldehyde-containing medium for 10 min at 37°C, and then they were sonicated to make soluble chromatin with DNA fragments between 200 and 1,000 bp.
3. The antibodies against protein tags such as Flag (M2, Sigma-Aldrich), c-myc (Santa Cruz Biotechnology, Santa Cruz, CA) or trimethylated H3-K4, acetylated histone H3 (Upstate Biotechnology, Lake Placid, NY) or transcription factors such as Sp1 (Santa Cruz Biotechnology) are used to precipitate DNA fragments bound by their corresponding elements.
4. The protein–DNA complex is collected with protein A- or G-Sepharose beads (Upstate Biotechnology), eluted, and reverse cross-linked.
5. After a treatment with Protease K (Sigma-Aldrich), the samples are extracted with phenol-chloroform and precipitated with ethanol.
6. The recovered DNA is resuspended in TE buffer and used for PCR amplification.

3.2.2. Development of Polyclonal Sera against Synthetic Peptides and Affinity Purification

This method is applied for H3 acetylation modification at lysine residues that undergo acetylation and phosphorylation *(13)*.

1. Peptides ARTKQTARK(Ac)STGGK(Ac)APRKQLC and ARTKQTARS(PO4)TGGKAPKQLC are selected based on their clinical potential and used to raised polyclonal antisera.
2. Very small amount of antigen (0.5 µg) is used for three times immunization keeping the interval 1 week apart

3. Specific IgG is purified from rabbit sera.
4. Peptide is coupled to PBS washed gel beads soaked in 1 M Tris-HCl, pH 8.8 (Pierce Chemical).
5. Free peptides are removed by PBS-Tween 20 and PBS washing.
6. IgG was eluted with glycine buffer and neutralized with Tris buffer.

3.2.3. Histone Methylation Assay

Because histone methylation is the most studied modification of histones, its assay method is described below.

1. Methylation binding domain protein (MBD1) is immunoprecipitated by IMG-306 antibodies (Imgenex) from the nuclear extract containing 50 µg of total nuclear proteins.
2. Protein complexes bound to protein G-Sepharose are extensively washed with NE buffer (20 mM HEPES, pH 7.0, 10 mM KCl, 1 mM $MgCl_2$, 0.5 mM DTT, 0.1% Triton, and 200 mM NaCl) and used for in vitro histone methylation assays using 0.5 µg recombinant histones or 1 µg recombinant histone H3 tails as substrates and 250 nCi/ml *S*-adenosyl-[methyl-^{3}H)-L-methionine.
3. The products are separated on 12% SDS-PAGE.
4. The gels are stained with Coomassie, soaked in Amplify reagent (GE Healthcare), dried, and exposed to BioMax MS film in BioMax Transcreen LE cassettes (Eastman Kodak, Rochester, NY) (*14*).

3.2.4. Immunoblotting of Histones

1. Cytoplasmic and nuclear extracts are prepared using NE-PER cytoplasmic extraction reagent (Pierce Chemical), according to the manufacturer's instructions.
2. For analysis of histone H3 modifications, cells are trypsinized at 37°C for 45–60 s, washed twice with ice-cold PBS, and resuspended in 15 µl of chilled lysis buffer per 10^6 cells (150 mM NaCl, 0.5% NP-40, 50 mM Tris, pH 8.0, 10 mM sodium butyrate, 1 mM phenylmethylsulfonyl fluoride, 0.5 mg/ml benzamidine, 2 µg/ml aprotinin, and 2 µg/ml leupeptin).
3. To ensure efficient cell lysis, samples are incubated on ice for 30 min, before centrifugation at 13,000 rpm at 4°C.
4. The supernatant is removed, and the insoluble pellet was resuspended in 4× strength Laemmli's buffer (15 µl/10^6 cells), boiled for 10 min, and passed through a 16-gauge needle.
5. Samples are stored at –70°C and reboiled before resolution by 15% SDS-PAGE.
6. Resolved proteins are electroblotted onto nitrocellulose membrane for 1 h at 25 V using the semidry transfer system (Bio-Rad, Hercules, CA).

7. Antibodies used to detect specific posttranslational modifications on histone H3 are as follows: anti-acetyl K9 antibody (Upstate Biotechnology), anti-acetyl K14 antibody (Upstate Biotechnology), anti-acetyl K18 antibody (Cell Signaling Technology, Beverly, MA), anti-acetyl K23 antibody (Cell Signaling Technology), and anti-phospho S10 antibody (Cell Signaling Technology).
8. Other antibodies used are DO-1 (Oncogene, against p53), 636 (Santa Cruz Biotechnology, against lamin A/C), and C20 (Santa Cruz Biotechnology, against p300).
9. Visualization of bound antibody is by enhanced chemiluminescence (Roche Diagnostics, Indianapolis, IN).
10. The intensity of bands is quantitated by the AlphaImager 1200 system (Alpha Innotech, San Leandro, CA) and AlphaImager 1200i 3.3b software by using scanned images with signal outputs within the linear range.

Histone modifications regulate epigenetic gene regulation in cancer cells and may serve as novel cancer diagnostic markers. Histones have four basic subunits, and in native state they exist as an octamer. DNA winds around this octamer. Acidic histones neutralize DNA charge and maintain chromatin stability. The octamer histone binding is independent of surrounding DNA sequence and N-terminal region of histones is subject to phosphorylation, acetylation, and other modifications. Multiple histone modifications can take place within a short stretch of amino acids of histone tails. These modifications regulate transcription, DNA replication and DNA repair, and thus are part of the transformation process.

Modification of histones may occur in large regions of chromatin, including coding and nonpromoter sequences, termed global histone modifications.

Global histone modifications have been suggested to play a role in cancer recurrence. Two enzymes are important in histone modifications: the histone acetyltransferase (HATs) adds the acetyl group to histones whereas the histone deacetylases (HDACs) remove the acetyl group. The opposing activities of HATs and HDACs tightly regulate gene expression through chromatin modifications. HATs activity results in chromatin relaxation, whereas HDACs make chromatin compact. Recently, histones have been investigated for use in cancer diagnosis and follow-up of treatments. A number of modifications in the tail region of histones have been reported, such as acetylation, deacetylation, phosphorylation, poly-ADP ribosylation, methylation, ubiquitination, sumoylation, carbonylation, citrullination, and glycosylation. It is important to understand the significance of these modifications, because they are not only important in cancer diagnosis but also to understanding the histone–DNA

interaction, nucleosome-nucleosome interaction, and interaction of nonhistone proteins with chromatin.

Histone modifications may produce two opposite effects at the level of transcription by either increasing or decreasing gene transcription. It has been proposed that a histone modification cassette exists that controls histone activities. In Section 3.2.4, we have described different types of histone modifications with relevant clinical information.

Acetylation is mediated by HATs where transfer of an acetyl group occurs from acetyl-coenzyme A to the epsilon-amino group of specific lysine side chain. The most common free amino terminal ends of lysine that undergo acetylation are lysine 9, 14, 18, and 23 in histone H3; and 5, 8, 12, and 16 of lysine in histone H4. The negatively charged DNA that generally interacts with histones is neutralized because acetylation interacts with the epsilon amino group of lysine. Furthermore, acetylated histones provide binding sites for different transcription factors. Modifications of lysine residues provide substances for generating specific antibodies that can be used for diagnosis of cancer. Several HATs have been identified (http://www.chromatin.us/hathome.html). HDACs have been categorized in three major classes: class I (HDACs 1, 2, 3, and 18), class II (HDACs 4, 5, 6, and 9), and class III (SIR2 family of NADP-dependent HDAC). HDACs deacetylase all four histones and sometimes act on nonhistone proteins as well.

The enzyme histone methyl transferase transfers the methyl group from *S*-adenosylmethionine to lysine or arginine of histones H3 and H4. Most of the time, lysine 4, 9, and 27 of H3 and lysine 20 of H4 get methylated. There are three types of methyltransferases: arginine methyl transferase (involved in maintenance of chromatin modifications); SET domain-containing methyl transferase, which induces ectopic heterochromatin; and Dot 1-like methyl transferase (involved in cell proliferation). Many transcription factors that interact with the promoter and histones get activated by phosphorylation. Aurora kinase is involved in histone H3 phosphorylation at serine 10 and 28 positions. Different forms of aurora kinase have different efficiencies of phosphorylation. Two other kinases that phosphorylate histones are Mst1 and RSK2 kinases. Mst1 has a preference for histone H2B, whereas RSK2 does not show specific preferences. Conversely, Histone H2B gets ubiquitinated by Rad6, which is an ubiquitinin-conjugating enzyme involved in DNA damage repair and meiosis. For transcription activity, sumoylation of histones occurs. The role of Ubc9 in sumoylation affecting normal mitosis and recovery from DNA damage is being investgated in different laboratories.

A few examples where histone modifications have been used in different tumor types follow. Dysplasia and neoplasia are difficult to distinguish by histopathologic analyses in cervical

cancer. However, nuclear phosphatase staining of histone H3 distinguishes these two stages of cancer development. Cytologic smears from case and control squamous epithelial cells were found positive with phosphorylated H3 antibodies. Success in the early detection and treatment of intraepithelial neoplasia has been implicated in protecting patients from developing invasive cervical cancer. Recently, the mortality due to cervical cancer decreased due to excellent screening programs using early marker detection.

Histone H3 methylation at lysine 4 and 9 has been studied in bladder cancer to investigate the effect of this modification on cancer-related gene expression. In one study, epigenetic modifications were followed in three human breast cancer cell lines MCF-7, MDA-MB-231, and MDA-MB-231(S30) representing different stages of human breast cancer. Decreased trimethylation of lysine 20 of histone H4, and hyperacetylation of histone H4, was observed in these cell lines. This dramatic decrease in trimethylation of lysine 20 of histone H4, especially in MDA-MB-231 cells, was accompanied by diminished expression of Suv4-20h2 histone methyltransferase. Thus, differential expression of histone modifications could be interpreted in cell lines as representing the aggressiveness of the tumor or different stages of tumor development. Such epigenetic dysregulation may contribute to and may be indicative of the formation of a more aggressive tumor phenotype during tumor progression. In another study, one of the HDACs, namely, HDAC6, was found to contribute to responsiveness to estrogen treatment.

In head and neck cancer, increased expression of *RAR* was observed. This increased expression was not due to methylation changes, but to increased acetylation of histone H3 at lysine 9. The expression of the receptor showed a correlation with the cell density.

Colon cancer diagnosis involves determining the ratio of methylation versus acetylation at the lysine 9 position of histone H3. This ratio also helps in distinguishing normal versus cancer cells. A high-throughput chip-on-chip assay has been developed for clinical application of histone modifications in cancer diagnosis. Hypoacetylation of histone H3 is shown to be useful, along with methylation markers in diagnosis of colorectal cancer. Histone H4 acetylation and hypermethylation of the fragile histidine triad gene are useful markers in following esophageal cancer development. H4 acetylation has been shown to be inversely proportional to the depth of cancer invasion and pathological stage. These markers also have been used for cancer prognosis.

Compared with other modifications, ubiquitination seems to dominate in lung cancer. Ubiquitination of H2A and H2B at different amino acid sites were studied in lung cancer cells exposed to nickel. Class II HDACs were found to be associated with

poor prognosis of lung cancer. Hypoacetylation of histones H3 and H4 and loss of trimethylation at lysine 4 of histone H3 was observed in five ovarian epithelial and carcinoma cell lines (human "immortalized" ovarian surface epithelium (HIO)-117, HIO-114, A2780, SKOV3, and ES2). Altered expression of GATA factors plays a crucial role in ovarian carcinogenesis. In the same study, lysine K9 dimethylation of histone H3 and HP1 association were not observed, excluding reorganization of GATA genes into heterochromatic structures. Results were confirmed using appropriate inhibitors of acetylation and methylation.

Like other cancers, methylation and histone modifications also have been reported in gastric cancer. In one study, hypoacetylation of histones H3 and H4 was observed in the *p21* promoter region of the gene in approximately half of the samples examined. Acetylation was correlated with advanced tumor stage, tumor invasion, and lymph node metastasis.

Epigenetic changes, including both methylation of promoter and modification of histone, occur quite early during cancer development. In prostate cancer diagnosis and recurrence, histone modifications have been used. Acetylation and methylation of different histones result in different outcomes and has a potential in prediction of clinical outcome and disease recurrence. Studying the histone modification profile also can help us understand different clinical outcomes.

Loss of monoacetylated and trimethylated forms of histone H4 in skin cancer has been observed early and being accumulated during cancer progression. Alteration at the acetylated Lys16 and trimethylated Lys20 residues of histone H4 were observed in cancer cells. Although validation has to be done by different groups, it seems that the global loss of monoacetylation and trimethylation of histone H4 is a common hallmark of human tumor cells.

The posttranslational modification of the core histones is critical to the regulation of chromatin structure. Recently, quantitative proteomic assays have been developed to check profiling of histone modifications, and it is possible that in the coming years these assays will be useable in clinics. Generally, it is much easier to study DNA modifications than protein modifications. In epigenetic regulation, histone modifications occur before DNA methylation. Because it is always to our advantage to diagnose cancer early, studying histone modifications has an advantage. Antibody based assays to detect histone modifications have been developed. A hydrophilic interaction liquid chromatographic separation method in combination with mass spectrometric analysis also has been used to follow the alterations in histone H4 methylation/acetylation status and the interplay between H4 methylation and acetylation during differentiation of erythroleukemia cells and study modifications that affect the chromatin structure.

Traditional methods use immunoassay techniques to determine the extent and site of posttranslational modification to detect histone modifications. Although these methods are sensitive, they require site-specific antibodies and knowing that a large number of modifications occur in histones, several antibodies would be available. Another approach that seems promising involves the application of reverse-phase high-pressure liquid chromatography and mass spectrometry (LC-MS) to analyze global modification levels of core histones. Relative to other methods, this method is fast, sensitive, and easily automatable. Furthermore, the LC-MS approach provides the global patterns of modification for all four core histones in a single experiment. The characterization of changes in histone modification in acute myeloid leukemia (AML) has been accomplished using LC-MS. Determination of dose-dependent changes in the distribution of modified core histones have been observed, and results have been validated in primary leukemia cells from patients with refractory or relapsed AML or chronic lymphocytic leukemia. For example, clinically significant higher levels of histone H4 were detected in patients at intervals of 4 and 24 h after treatment with inhibitors. Attempts should be made to establish a panel of antibodies against histone modifications that could discriminate different stages of tumor development. This would help in screening patient sera on microarrays with a combination of antibodies. Ultimately, samples from populations with cancer, populations at high risk of developing cancer, or both would be identified using approaches described in Sections 3.2.3 and 3.2.4. It also would help in the identification of associated etiologic factors that contribute to the initiation and progression of cancer.

4. Chromatin Modeling

4.1. Materials for Determining Chromatin Modeling

1. PC8-Phenol/Chloroform/Isoamyl Alcohol, 25:24:1 (v/v), molecular biology grade (catalog no. 516726, EMD Biosciences, San Diego, CA).
2. I-SAGE Long kit (catalog no. T5000-03, Invitrogen).
3. AscI biotinylated linker (gel-purified), sense 5′-biotin-TTTGCAGAGGTTCGTAATCGAGTTGGGTGG-3′ and antisense 3′-CGTCTCCAAGCATTAGCTCAACCCACCGCGC-phos-5′ (Integrated DNA Technologies, Coralville, IA).
4. Rapid DNA ligation kit (catalog no. 1 635 379, Roche Diagnsotics).
5. Streptavidin (catalog no. S4762, Sigma-Aldrich).

6. Dynabeads M-280 streptavidin (catalog no. 112-05D, Invitrogen).
7. T4 DNA ligase HC and 5× ligase buffer (catalog no. 15224-041, Invitrogen).
8. Platinum Taq DNA Polymerase (catalog no. 10966-034, Invitrogen).
9. Sephadex G25 (catalog no. 17057202, GE Healthcare).
10. Siliconized tubes (catalog no. 12450, Ambion, Austin, TX).
11. Magnetic stand (catalog no. R670-01, Invitrogen).
12. Phase-lock gel tube light (catalog no. 0032 007.961, Eppendorf North American, New York, NY).
13. Costar Spin-X centrifuge tube filters (catalog no. 8161; Corning Life Sciences, Acton, MA).
14. I-SAGE Long kit (catalog no. T5000-03, Invitrogen); the components included in the kit are as follows: SNAP Columns and Collection Tubes (can be replaced by Costar Spin-X centrifuge tube filters).
15. Vertical electrophoresis apparatus (catalog no. 16X14 09-518-157C, Fisher Scientific, Pittsburgh, PA).
16. Gene pulser cuvettes (catalog no. 165-2089, Bio-Rad).
17. Nunc plates (245 × 245 mm) (catalog no. 12565224; Fisher Scientific).

4.2. Methods for Determining Chromatin Modeling

4.2.1. Chip-on-Chip

For multiple sample analysis Agilent Technologies (Santa Clara, CA) has developed microfluidic Lab-on-a-Chip technology.

1. This technology uses a network of channels and wells that are etched onto glass or polymer chips to build mini-labs.
2. Pressure or electrokinetic forces move picoliter volumes in finely controlled manner through the channels.
3. Lab-on-a-Chip enables sample handling, mixing, dilution, electrophoresis and chromatographic separation, staining and detection on single integrated systems.
4. The main advantages of Lab-on-a-Chip are ease-of-use, speed of analysis, low sample, and reagent consumption and high reproducibility due to standardization and automation.
5. Agilent has introduced the first commercial CpG Island microarrays for studying DNA methylation.
6. Recently, microarrays have evolved as a powerful analysis tool for genome-wide measurement of DNA methylation changes.

4.2.2. Tiling Array

1. Epigenomic changes play a crucial role in transcriptional regulation, because aberrant DNA methylation provokes silencing of tumor suppressor genes and chromosomal instability in cancer.
2. Previous studies have been limited to specific genes, promoters, and CpG islands (CGIs).
3. A tiling array analysis, combined with methylated DNA immunoprecipitation, has enabled global analysis of DNA methylation patterns. Using this approach, >700 methylated sites in genomic DNA from human colorectal cancer cell lines within the ENCODE region, more than half of which were overlapped with CGIs have been identified.
4. These methylated sites were enriched with transcription regulatory elements, such as non-coding upstream sequences and 5 kb within both ends of mRNAs. Histone acetylation patterns in HOXA cluster also was evaluated, and a reciprocal relationship between DNA methylation and histone H3 and H4 acetylation was observed. Tiling array analysis is easily scalable to the entire genome and will be an indispensable tool for the regulome analysis.

Chromatin exists in euchromatin or heterochromatin states. Transcriptionally active chromatin is euchromatin and during this state several transcription factors bind to the promoter region *(15)*. Chromosomal rearrangements also give rise to altered chromatin compactation. These rearrangements can be detected using fluorescence in situ hybridization and three-dimensional microscopy. One such example is thyroid cancer where the locations of five different loci in this region were mapped *(15)*. Another example where chromatin modeling was used to understand carcinogenesis is regulation of multidrug resistance (*MDR*) gene *(16)*. For long time, the origin of drug-resistant cells in human cancers has been a major point of discussion. Both genetic and epigenetic mechanisms are involved in developing multiple drug resistance via *MDR1* gene. Drug resistance involves initiation of transcription at a site 112 kb upstream of the ABCB1 proximal promoter (P1) in the drug-resistant cells mediated by chromatin-remodeling process characterized by an increase in acetylated histone H3 within a 968-bp region 5′ of the ABCB1 upstream promoter. Various theories have been put forwarded for different drug exposure; however, Chen's model provides evidence for a major genomic alteration that changes the chromatin structure of the ABCB1 upstream promoter via acetylation in MES-SA cells *(16)*. Another example where chromatin modeling was useful in disease diagnosis is leukemia. In AML, chromosomal translocations result in formation of chimeric transcription factors, such as PML/RARα, PLZF/RARα, and AML-1/ETO, of which the components are involved in regulation of transcription by chromatin modeling *(17)*.

The tumor suppressor *p53* has an essential role in preventing genomic instability and tumor development in mammals. This effect is achieved in part by controlling the expression of its target genes in response to various cellular stresses. Recently, *p53*-mediated histone modifications have been shown to play roles in the transcriptional regulation; however, there still exists lack of comprehensive understanding of the association of p53 with histone modification in the human genome. By combining chromatin immunoprecipitation with oligonucleotide tiling arrays, p53-binding sites (p53-BS) in HCT116 human colon carcinoma cells, together with acetylated H3 (Lys9/14), acetylated H4 (Lys5/8/12/16), di-methylated H3 (Lys4), and tri-methylated H3 (Lys4) sites along the human genome have been mapped. Thirty-seven p53-BS, all of which were confirmed by quantitative PCR to have enrichment ratios of at least 2-fold against the background were observed. The p53-binding motif was found in most of the p53-BS and histone acetylation was observed in more than half of the p53-BS. On the contrary, only a small portion of the p53-BS was accompanied by histone methylation. Combined with distinctive genomic distribution of the modified histones, this information provided novel insights into how p53-binding in association with the histone modification coordinates transcriptional regulation.

5. Concluding Remarks and Future Directions

Cancer epigenetics is evolving rapidly on several fronts, including cancer epidemiology. Due to the development of high-throughput methods of detecting epigenetic modifications, it is now possible to identify one population from others, as discussed here for colon cancer. Epigenetic markers are being used in clinics for classification of few cancer types. These markers also are useful for assessment of risk of developing cancer.

For several histone modifiers, we do not know the mechanism of action. Technological advancements in microarray technologies and proteomics may shed some light on this aspect. Genome-wide mapping approaches of histones provide new opportunities to decipher histone code. Integration of datasets and improved technologies will provide future directions. The major challenge in the area of cancer diagnosis is validation of markers that have been studied by one or few groups. As always, this will be a very big project in which the collaborative efforts of biochemists, proteomic and genomic experts, clinicians, and molecular biologists will be needed.

References

1. Jones, P. A., and Baylin, S. B. (2007) The epigenomics of cancer. *Cell* **128,** 683–692.
2. Jones, P. A., and Martienssen, R. (2005) A blueprint for a Human Epigenome Project: the AACR Human Epigenome Workshop. *Cancer Res* **65,** 11241–11246.
3. Marjoram, P., Chang, J., Laird, P. W., and Siegmund, K. D. (2006) Cluster analysis for DNA methylation profiles having a detection threshold. *BMC bioinformatics* **7,** 361.
4. Siegmund, K. D., Levine, A. J., Chang, J., and Laird, P. W. (2006) Modeling exposures for DNA methylation profiles. *Cancer Epidemiol Biomarkers Prev* **15,** 567–572.
5. Verma, M., Seminara, D., Arena, F. J., John, C., Iwamoto, K., and Hartmuller, V. (2006) Genetic and epigenetic biomarkers in cancer: improving diagnosis, risk assessment, and disease stratification. *Mol Diagn Ther* **10,** 1–15.
6. Eads, C. A., Danenberg, K. D., Kawakami, K., Saltz, L. B., Blake, C., Shibata, D., Danenberg, P. V., and Laird, P. W. (2000) MethyLight: a high-throughput assay to measure DNA methylation. *Nucleic Acids Res* **28,** E32.
7. Fiegl, H., Millinger, S., Goebel, G., Muller-Holzner, E., Marth, C., Laird, P. W., and Widschwendter, M. (2006) Breast cancer DNA methylation profiles in cancer cells and tumor stroma: association with HER-2/neu status in primary breast cancer. *Cancer Res* **66,** 29–33.
8. Eads, C. A., Lord, R. V., Wickramasinghe, K., Long, T. I., Kurumboor, S. K., Bernstein, L., Peters, J. H., DeMeester, S. R., DeMeester, T. R., Skinner, K. A., and Laird, P. W. (2001) Epigenetic patterns in the progression of esophageal adenocarcinoma. *Cancer Res* **61,** 3410–3418.
9. Genereux, D. P., Miner, B. E., Bergstrom, C. T., and Laird, C. D. (2005) A population-epigenetic model to infer site-specific methylation rates from double-stranded DNA methylation patterns. *Proc Natl Acad Sci U S A* **102,** 5802-5807.
10. Widschwendter, M., Fiegl, H., Egle, D., Mueller-Holzner, E., Spizzo, G., Marth, C., Weisenberger, D. J., Campan, M., Young, J., Jacobs, I., and Laird, P. W. (2007) Epigenetic stem cell signature in cancer. *Nat Genet* **39,** 157–158.
11. Burden, A. F., Manley, N. C., Clark, A. D., Gartler, S. M., Laird, C. D., and Hansen, R. S. (2005) Hemimethylation and non-CpG methylation levels in a promoter region of human LINE-1 (L1) repeated elements. *J Biol Chem* **280,** 14413–14419.
12. Stoger, R., Kajimura, T. M., Brown, W. T., and Laird, C. D. (1997) Epigenetic variation illustrated by DNA methylation patterns of the fragile-X gene FMR1. *Hum Mol Genet* **6,** 1791–1801.
13. Anton, M., Horky, M., Kuchtickova, S., Vojtesek, B., and Blaha, O. (2004) Immunohistochemical detection of acetylation and phosphorylation of histone H3 in cervical smears. *Ceska Gynekol* **69,** 3–6.
14. Allison, S. J., and Milner, J. (2003) Loss of p53 has site-specific effects on histone H3 modification, including serine 10 phosphorylation important for maintenance of ploidy. *Cancer Res* **63,** 6674–6679.
15. Gandhi, M., Medvedovic, M., Stringer, J. R., and Nikiforov, Y. E. (2006) Interphase chromosome folding determines spatial proximity of genes participating in carcinogenic RET/PTC rearrangements. *Oncogene* **25,** 2360–2366.
16. Chen, K. G., Wang, Y. C., Schaner, M. E., Francisco, B., Duran, G. E., Juric, D., Huff, L. M., Padilla-Nash, H., Ried, T., Fojo, T., and Sikic, B. I. (2005) Genetic and epigenetic modeling of the origins of multidrug-resistant cells in a human sarcoma cell line. *Cancer Res* **65,** 9388–9397.
17. Puccetti, E., Obradovic, D., Beissert, T., Bianchini, A., Washburn, B., Chiaradonna, F., Boehrer, S., Hoelzer, D., Ottmann, O. G., Pelicci, P. G., Nervi, C., and Ruthardt, M. (2002) AML-associated translocation products block vitamin D(3)-induced differentiation by sequestering the vitamin D(3) receptor. *Cancer Res* **62,** 7050–7058.

Chapter 15

Mitochondrial DNA Polymorphism and Risk of Cancer

Keshav K. Singh and Mariola Kulawiec

Summary

ATP (energy production) production is not the only function of the mitochondria. Mitochondria perform multiple cellular functions. Among others, these functions include control of cell death, growth, development, integration of signals from mitochondria to nucleus and nucleus to mitochondria, and various metabolic pathways. Although defects in mitochondrial function are most commonly associated with bioenergetic deficiencies, our studies demonstrate that mitochondrial defects lead to genome instability in the nuclear DNA, resistance to apoptosis and induction of NADPH oxidase, a designated producer of reactive oxygen species. These transformations in cellular phenotype are known contributors to the development of tumors in humans. Consistent with the role of mitochondria in carcinogenesis, studies in the past few years have described an increased risk of cancers associated with specific mitochondrial DNA (mtDNA) polymorphism among various different haplogroups in human population. However, molecular mechanisms underlying increased risk of cancer due to specific mtDNA polymorphisms is currently lacking. It is likely that mtDNA polymorphisms in mitochondrial genes involved in electron transport chain and oxidative phosphorylation result in increased oxidative stress and hypermutagenesis of mitochondrial as well as nuclear DNA. We suggest that in studies relating to cancer epidemiology, the significance of a particular mtDNA polymorphism(s) should be analyzed together with other polymorphisms in mtDNA and in nuclear DNA.

Key words: Apoptosis, mitochondria, oxidative phosphorylation, polymorphism, reactive oxygen species.

1. Introduction

Mitochondria perform multiple cellular functions. Mitochondria are the predominant source of energy (ATP) production and reactive oxygen species (ROS) in the cell. ROS are generated in the mitochondria as a by-product of oxidative phosphorylation (OXPHOS) involved in ATP production. Mitochondria contain their own genetic material (mitochondrial DNA, mtDNA) which

Mukesh Verma (ed.), *Methods in Molecular Biology, Cancer Epidemiology, vol. 471*
© 2009 Humana Press, a part of Springer Science + Business Media, Totowa, NJ
Book doi: 10.1007/978-1-59745-416-2

encodes proteins that are an integral part of various OXPHOS complexes *(1,2)*. It is estimated that 90% of cellular oxygen is consumed in the mitochondria and about 2–4% of that oxygen is converted to ROS. This amounts to approximately 10^7 ROS molecules/mitochondrion every day *(1,2)*. Consistent with these observations, studies have indicated somatic mtDNA mutations in almost all cancers examined to date. The mutations found in tumors cells reinforce the suggestion made by Warburg in 1930 *(3)*. Warburg hypothesized that mitochondrial abnormalities were implicated in development of cancers. Since then, many mitochondrial abnormalities in cancers, both at genetic and metabolic levels, have been reported. In this chapter, we intend to discuss abnormalities at the mtDNA level. With regard to mitochondrial metabolic and other related abnormalities in cancers, see recent reviews *(1,2)*.

2. Mitochondrial Genome

The mitochondrial genome is a 16.6-kb closed circular, double-helical molecule. The mitochondrial genome is inherited only through the mother. The mtDNA encodes two rRNAs, 22 tRNAs, and 13 proteins (**Fig. 15.1**; *[4]*). Each of these proteins is a subunit of one of four respiratory enzyme complexes localized into the mitochondria. They include seven subunits of respiratory enzyme complex I, one subunit of complex III, three subunits of complex IV, and two subunits of complex V. All other (~1,500) mitochondrial proteins, including those involved in the replication, transcription, and translation of mtDNA, are encoded by nuclear genes, and they are targeted to the mitochondrion by a specific transport system *(5)*. Although mtDNA represents <1% of total cellular DNA, its gene products are essential for normal cell function. Unlike nuclear DNA, mammalian mtDNA contains no introns, has no protective histones, and is exposed to deleterious reactive oxygen species generated by OXPHOS. In addition, replication of mtDNA is error prone *(4)*. The accumulation of mutations in mtDNA is approximately 10-fold greater than that in nuclear DNA *(6,7)*. This feature of mtDNA makes it a useful tool for studying human migration around the globe and identifying specific human populations.

A haplogroup is defined on the basis of a particular mutation that is well established and widely distributed among individual of the populations. These haplogroups can be further divided into haplotypes, generally based on restriction fragment length polymorphism (RFLP). RFLP are a result of restriction enzymes that cut DNA sequences at a particular point, priming them for electrophoresis and allowing a comparison to be made

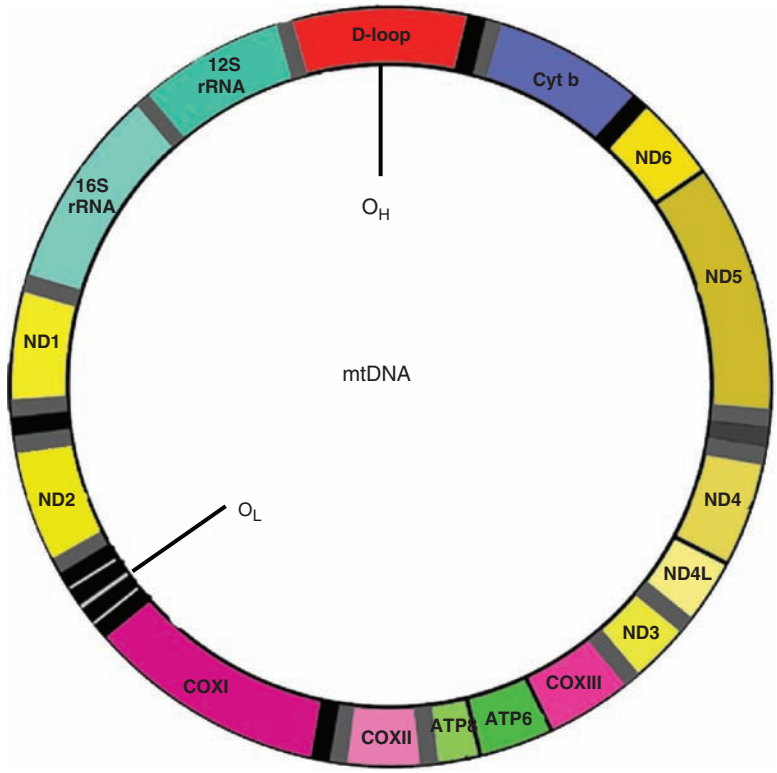


Fig. 15.1. mtDNA map.

showing a pattern among individuals based on the length of the DNA fragments *(8)*. More than 2 dozen mtDNA haplogroups are known among human populations around the world. **Figure 15.2** describes some of the major mitochondrial haplogroups identified throughout the world. **Figure 15.2** also reveals the RFLPs used for haplotype analyses. However, haplogroups and subgroups are best defined by mtDNA sequencing.

3. Mitochondrial Polymorphism and Risk of Cancer

mtDNA mutations causing defects in mitochondrial respiratory enzyme complexes are thought to increase risk of cancer. In support of this hypothesis, we have demonstrated that defects in mitochondrial respiratory chain activity lead to 1) overexpression of superoxide producing NADPH oxidase *(9)*; 2) increased damage and hypermutagenesis in the nuclear DNA *(10–13)*; and 3) resistance to apoptosis *(14)*. These transformations in cellular phenotype are known to contribute to the development of cancer.

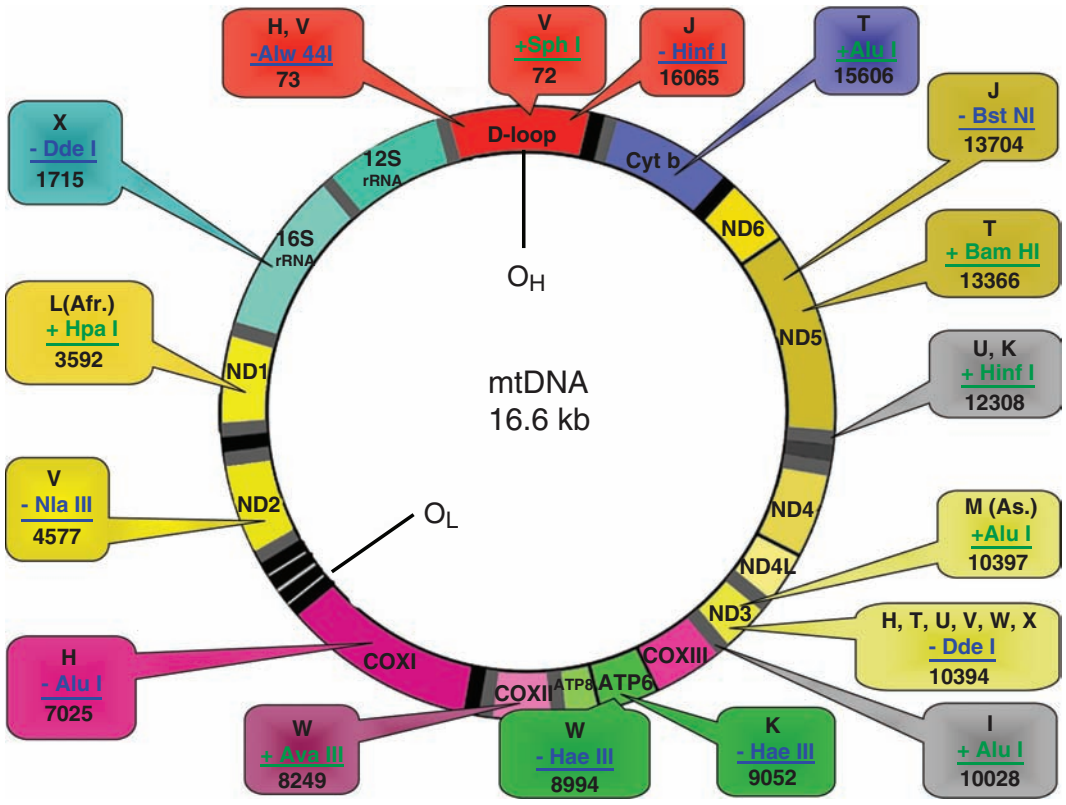


Fig. 15.2. Haplotype analysis by RFLP: summary of various restriction enzymes used to define haplotype.

Studies in the past few years have described the risk of cancers associated with various mtDNA haplogroups in the human population. Tanaka's group analyzed a population with 1,503 autopsied cases (696 male, 807 female) at Tokyo Metropolitan Geriatric Hospital registered. The genotypes for 149 polymorphisms in the coding region of the mitochondrial genome were determined. The haplogroups, were classified into 30 haplotypes, i.e., F, B5, B4a, B4b, B4c, A, N9a, N9b, Y, M10+M11+M12, M7a, M7b2, M7c, M8+Z+C, G1, G2, M9, D5, D4a, D4b, D4d, D4e, D4g, D4h, D4j, D4k, D4k, D4l, D4m, and D4n. Multivariate logistic regression analysis was performed with adjustment for age, the prevalence of drinking, and that of smoking. Among 1,363 subjects whose data were available for presence or absence of cancer, age, gender, history of alcohol intake, history of smoking, and mitochondrial haplogroups, 819 subjects carried a pathologically confirmed cancer(s) at the time of autopsy. Subjects with the haplogroup M7b2 tended to have an increased risk for hemopoietic cancer ($P = 0.037$), having an odds ratio of 2.46 (95% confidence interval [CI] = 1.06–5.73).

Results also indicated that haplogroup M7b2 is a risk factor for leukemia *(15)*.

Nine main European haplotypes (H, I, J, K, T, U, V, W, and X) were analyzed in a series of patients with prostate and renal cancers studied by Booker et al. *(16)*. Using the A12308G substitution in $tRNA^{Leu2}$ (**Fig. 15.3**) as a marker of the mtDNA haplogroup U, it was found that patients carrying this haplogroup had an increased risk of renal (odds ratio [OR] = 2.52) and prostate cancer (OR = 1.95) *(16)*.

Canter's epidemiologic studies in breast cancer showed the G10398A substitution in ND3 was associated with an increased risk (OR = 1.6) in African-American women (**Fig. 15.3**; *(17)*). His group also reported one of the first cases of synergy among single-nucleotide polymorphisms. The T4216C substitution alone in ND1 conferred no increased risk. However, when the T4216C substitution was present with G10398A, the risk of breast cancer was increased (OR = 3.1; **Fig 15.3**). It has been hypothesized also that the prevalence of germline cyclooxygenase (COX) I mutations may explain the increased incidence of prostate cancer in African American men *(18)*.

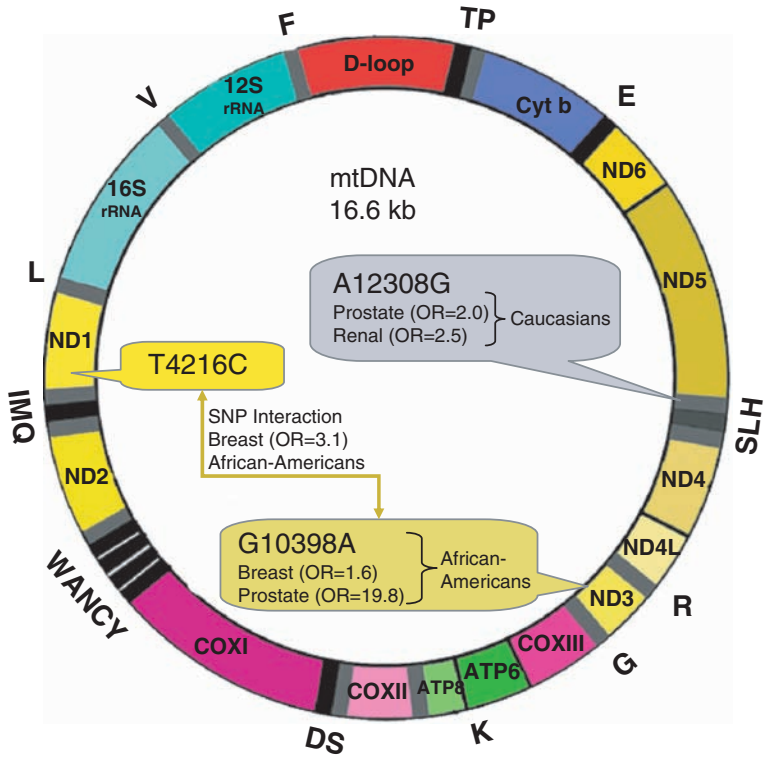


Fig. 15.3. mtDNA oncomap: summary of mtDNA mutations found in various tumors examined.

Darvishi et al. *(19)* analyzed mtG10398A (Ala→Thr) polymorphism in a haplotype constituting mtDNA haplogroup N and its sublineages for its contribution to higher risk for breast cancer. They analyzed approximately 1,000 complete human mtDNA sequences worldwide and collated information on 2,334 individuals belonging to 18 regions in India. They confirmed that mt10398A allele provides a risk toward breast cancer by using a case-control comparison study of 124 sporadic breast cancer patients and 273 controls. They also included 55 squamous cell carcinoma of esophagus and 163 controls, matched for age, ethnicity, and sex from north India. Their study also found that 10398A mtDNA polymorphism confers a higher risk for esophageal cancer. In a recent study, Li et al. *(62)* showed an association between mtDNA haplogroup D4a and D5a and increased risk of esophageal cancer in China.

Bai et al. *(20)* analyzed mtDNA polymorphism in European-American females and reported that A10398G (OR = 1.79) and T16519C (OR = 1.98) increase breast cancer risk. In contrast, T3197C (OR = 0.31) and G13708A (OR = 0.47) were found to decrease breast cancer risk.

Wang et al. *(21)* evaluated polymorphisms in mtDNA associated with increased risk of pancreatic cancer. They screened Caucasian cases and found no significant associations with pancreatic cancer. It is likely that polymorphism in other mitochondrial genes or in conjunction with nuclear genes involved in OXPHOS may contribute to pancreatic cancer.

In some of the above-mentioned studies, it is noteworthy that the potential influence of population stratification was not taken in to account. A population stratification of these studies, independent positive replication of these studies, or both are needed to provide convincing evidence for the association of mtDNA polymorphism and risk for cancer.

4. Mitochondrial Polymorphism and Risk of Other Diseases

Recent studies in Japanese centenarians and super-centenarians (those ≥105 years) have demonstrated that specific mtDNA haplogroup that are associated with these age groups. Centenarians seem to be resistant against lifestyle-related diseases such as type 2 diabetes, myocardial infarction, and cerebrovascular infarction, as well as geriatric diseases such as Alzheimer's disease (AD) and Parkinson disease *(22)*.

Tanaka's group described that mtDNA mutations in these age groups provide protection from disease and confer longevity. Data from this study suggest that longevity is associated with

the D4a haplogroup, whereas noncentenarians mitochondrial haplogroups F and A are risk factors for diabetes. Furthermore, haplogroup N9a was found to be a protective against diabetes, especially in women. A large-scale association studies will have a major impact on the functional studies of mitochondrial haplogroups and on the elucidation of their contributions to longevity or pathogenesis of lifestyle-related diseases *(23)*.

Van der Walt *(24)* analyzed risk of AD associated with European mitochondrial haplogroups. Interestingly, they found that males in haplogroup U showed an increase in risk (OR = 2.30; 95% CI = 1.03–5.11; $P = 0.04$) of AD relative to the most common haplogroup H, whereas females demonstrated a significant decrease in risk with haplogroup U (OR = 0.44 ; 95% CI = 0.24–0.80; $P = 0.007$).

Van der Walt *(24)* also demonstrated a significant decrease in risk of PD among European haplogroup J versus individuals carrying the most common haplogroup, H. Furthermore, a specific mutation 10398G, that defines these two haplogroups, is strongly associated with this protective effect (OR = 0.53; 95% CI = 0.39–0.73; $P = 0.0001$). Interestingly, these studies suggest that a single mtDNA substitution may exert different effects in different disease states.

5. Mitochondrial DNA Mutations in Human Tumors

Mutations in mtDNA have been reported in all cancers examined to date *(1,2)*. A summary of mtDNA mutations in various tumors is presented in **Fig. 15.4**. Although these mutations are known to occur throughout the mitochondrial genome, the displacement loop (or D-loop) region has been shown to be a mutational "hot spot" in human cancer. The D-loop is a noncoding control region of mtDNA (np 16024–516) that houses *cis*-regulatory elements required for replication and transcription of the molecule. Thus, mtDNA mutations in this region might be expected to affect copy number and gene expression of the mitochondrial genome. Somatic mutations in the D-loop region have been observed in all tumors examined to date *(1,2)*. In a recent comprehensive study involving 54 hepatocellular carcinomas, 31 gastric, 31 lung, and 25 colorectal cancers, the incidence of somatic D-loop mutations in cancers of later stage was found to be higher that that of early stage cancers *(25)*. These findings suggest that instability in the D-loop region of mtDNA may be involved in carcinogenesis of human cancers.

In addition to the mtDNA mutations found in the D-loop region, deletions, point mutations, insertions, and duplications in other parts of the mitochondrial genome have been noted in

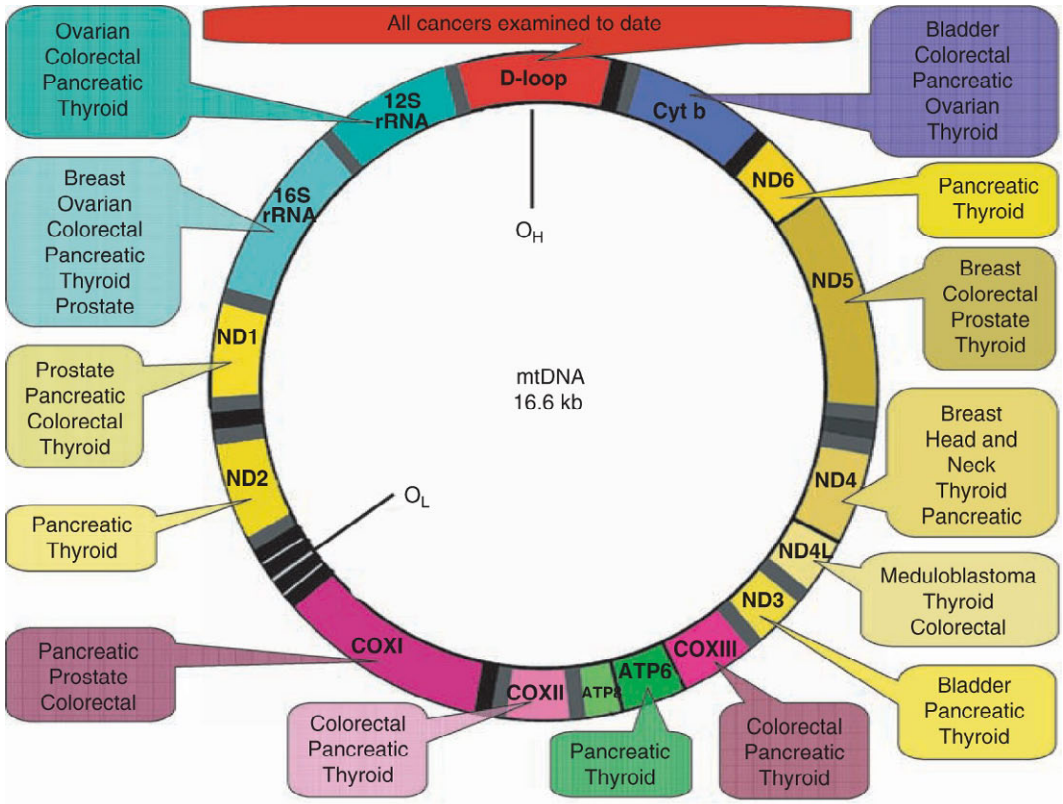


Fig. 15.4. mtDNA polymorphism associated with risk of cancers: G10398A in ND3 is associated with increased risk of breast (OR = 1.6) and prostate cancer (OR = 19.8) in African-Americans *(37)*. A12308G in $tRNA^{Leu2}$ is a marker of the mtDNA haplogroup U. This haplogroup is associated with increased risk of prostate (OR = 1.95) and renal cancer (OR = 2.52) in Caucasians. *, the T4216C substitution in ND1 conferred no increased risk in isolation. However, when the T4216C substitution was present with G10398A, synergy was observed, and the risk of breast cancer was increased (OR = 3.1) in African-Americans.

a variety of human cancers, including ovarian, thyroid, salivary, kidney, liver lung, colon, gastric brain, bladder, head and neck, prostate, and breast cancer, and leukemia *(26,18)*. For example, a 40-bp insertion localized in the COX I gene seems to be specific for renal cell oncocytoma *(27)*, and a deletion mutation resulting in the loss of mtDNA within NADH dehydrogenase subunit III is associated with renal carcinoma *(28)*. In a recent population based study involving 260 prostate cancer patients of European and African-American descent and 54 benign controls without cancer, the frequency of COX I missense mutations was found to be significantly higher in prostate cancer patients compared with the no-cancer controls (12 vs. 1.9%, respectively). At least some of these COX I sequence variants were thought to represent germ line mutations) *(18)*. The co-occurrence of somatic and germline mtDNA mutations in renal cell carcinoma also has been reported *(29)*.

Tumor-specific changes in mtDNA copy number also have been noted in human cancers. For example, the mtDNA content was found to be elevated in primary tumors of head and neck squamous cell carcinoma *(30)*, and to increase with histopathologic grade in premalignant and malignant head and neck lesions *(31)*. In addition, mtDNA copy number was shown to increase in papillary thyroid carcinomas *(32)* and during endometrial cancer development *(33)*. Conversely, it was reported that mtDNA content is reduced in 80% of breast tumors relative to normal controls. mtDNA depletion also occurs in gastric cancers *(32,33)* and during endometrial cancer development *(34)*. Conversely, it was reported that mtDNA content is reduced in 80% of breast tumors relative to normal controls. mtDNA depletion also occurs in breast cancers *(35)*. The actual mtDNA copy number in cancers might depend upon the specific site of mutation associated with that cancer. For example, mutations in the D-loop region, which controls mtDNA replication, would be expected to result in a decrease in copy number. Conversely, an increase in mtDNA copy number might occur as a compensatory response to mitochondrial dysfunction or due to mutations in nuclear genes involved in controlling mtDNA copy number indirectly.

Mitochondrial abnormalities resulting from mutation in mtDNA or copy number changes invokes mitochondria-to-nucleus retrograde response in human cells *(36)*. To identify proteins involved in retrograde response and their potential role in tumorigenesis, Kulawiec et al. carried out a comparative proteomic analysis using a parental cell line, a rho^0 cell line in which the mitochondrial genome was completely depleted (and the cells were therefore lacking all mtDNA-encoded protein subunits), and a cybrid cell line in which mtDNA was restored *(10,36)*. This comparative proteomic approach revealed marked changes in the cellular proteome, including quantitative changes in expression of several proteins in breast and ovarian tumors, which suggest that retrograde responsive genes may potentially function as tumor suppressor or oncogenes.

Transmitochondrial hybrid (cybrid) studies also suggest that mtDNA plays a key role in establishing the tumorigenic phenotype, maintaining the tumorigenic phenotype, or both. In a recent study, Singh et al. demonstrated that a rho^0 derivative of human osteosarcoma cells display increased tumorigenic phenotype as evidenced by increased anchorage independent growth compared with the parental cells *(10)*. Interestingly, the tumorigenic phenotype was restored to parental level by transfer of wild-type mitochondrial DNA to rho^0 cells. These studies suggest that intergenomic cross-talk between mitochondria and cancer plays an important role in tumorigenesis and that mitochondria-to-nucleus retrograde signaling may be an important

factor in restoration of the nontumorigenic phenotype. Indeed, additional studies by Singh's group demonstrate that retrograde mitochondria-to-nucleus signaling involves regulation of NADPH oxidase (Nox1) and that this enzyme is overexpressed in a majority of breast and ovarian tumors *(9)*. The family of Nox enzymes comprises seven structurally related homologs, Nox 1–5 and dual oxidase 1 and 2 *(37,38)*, and nuclear DNA encoded Nox 1 is a major source of endogenous ROS in the cell *(39)*. Until recently, these enzymes were known only to function as phagocytic respiratory burst oxidases that catalyze the NADPH-dependent reduction of molecular oxygen to superoxide, hydrogen peroxide, and other ROS, and to participate in host defense by killing or damaging invading microbes *(39)*.

Transmitochondrial cybrid studies also suggest the functional significance of mtDNA mutations. Cybrids harboring the ATP6T8993G mtDNA mutation in prostate cancer (PC3) cells were found to generate tumors that were 7 times larger than wild-type cybrids, which barely grew in mice *(8)*. Additionally, cybrids constructed using a common HeLa nucleus and mitochondria containing a point mutation in ATP synthase subunit 6 were conferred a growth advantage in early tumor stages after transplantation into nude mice. It was suggested that this growth advantage might possibly occur via prevention of apoptosis *(40)*. These studies support the hypothesis that mtDNA mutations might directly promote tumor growth in vivo.

6. Possible Mechanism of mtDNA-Mediated Carcinogenesis

Polymorphisms in human mtDNA may affect the efficiency of ETC and ROS production. This argument is supported by the fact that common mtDNA variants are indeed involved in ROS production *(41)*. Mitochondrial respiratory activity is associated with the generation of ROS. Under physiological conditions, a small fraction of reducing equivalents from complex I, complex III, or both of the mitochondrial electron transport chain may be transferred directly to molecular oxygen, generating the superoxide anion O_2^-. This ROS is converted to H_2O_2 by mitochondrial manganese superoxide dismutase. H_2O_2 can be converted to water by glutathione peroxidase or catalase, or it can acquire an additional electron from a reduced transition metal to generate the highly reactive hydroxyl radical ·OH. ROS generation has been shown to increase in mitochondria under conditions of excess electrons or as a result of respiratory enzyme complex inhibition. High levels of ROS, or oxidative stress, can not only mutagenize mtDNA but also induce hypermutagenesis and instability

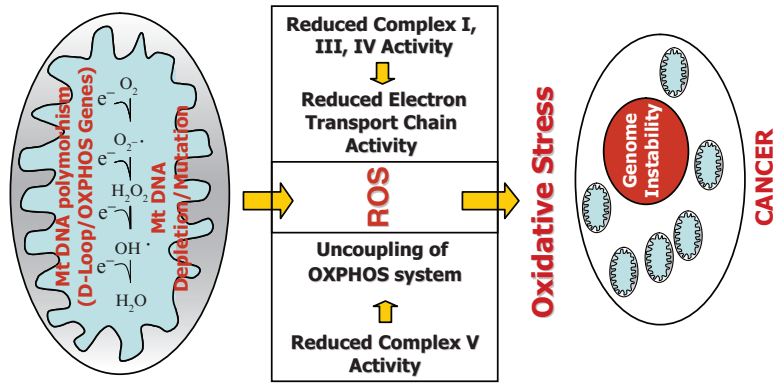


Fig. 15.5. Mechanisms of mitochondria-mediated mutagenesis and development of cancer (see text for details).

of nuclear (chromosomal) DNA (**Fig. 15.5**). These factors can contribute to development of cancer.

7. Future Prospectives

In past year, studies have been conducted that identify the role of mtDNA polymorphism of cancer. However, to fully understand the importance of mtDNA polymorphism and risk of cancer, it is necessary to 1) identify genetic and epigenetic changes in the nucleus associated with mtDNA polymorphisms; 2) identify D-loop region polymorphisms resulting in changes in mtDNA copy number; 3) identify nuclear genes controlling mtDNA copy number; 4) identify functional changes associated with mtDNA copy number; 5) identify genetic alteration in the nuclear DNA (nDNA) associated with mtDNA copy number changes; 5) identify epigenetic alteration in the nDNA associated with mtDNA copy number changes; and understand the role of synergy between different polymorphisms in mtDNA as well in nDNA. This is particularly important in view of constant cross-talk between the mitochondria and mitochondria and nucleus.

Acknowledgments

Studies in our laboratory were supported by National Institutes of Health grants R01 CA121904, 113655 and 116430 (to K.K.S.) and the New York State Department of Health (to M.K. and K.K.S.).

References

1. Modica-Napolitano, J. S., and Singh, K. K. (2004) Mitochondrial dysfunction in cancer. *Mitochondrion* **4**, 755–762.
2. Modica-Napolitano, J. S., and Singh, K. (2002) Mitochondria as targets for detection and treatment of cancer. *Expert Rev Mol Med* **4**, 1–19.
3. Warburg, O. (1930) *Metabolism of Tumors*, Arnold Constable, London, UK.
4. Singh, K. K. (ed.) (1998) *Mitochondrial DNA Mutations in Aging, Disease, and Cancer*. Springer, New York.
5. Schatz, G. (1996) The protein import system of mitochondria. *J Biol Chem* **271**, 31763–31766.
6. Grossman, L. I., and Shoubridge, E. A. (1996) Mitochondrial genetics and human disease. *Bioessays* **18**, 983–991.
7. Johns, D. R. (1995) Seminars in medicine of the Beth Israel Hospital, Boston: mitochondrial DNA and disease. *N Engl J Med* **333**, 638–644.
8. Malhi, R. S., Mortensen, H. M., Eshleman, J. A., Kemp, B. M., Lorenz, J. G., Kaestle, F. A., Johnson, J. R., Gorodezky, C., and Smith, D. G. (2003) Native American mtDNA prehistory in the American Southwest. *Am J Phys Anthropol* **120(2)**, 108–24.
9. Desouki, M. M., Kulawiec, M., Bansal, S., Das, G. M., and Singh, K. K. (2005) Cross-talk between mitochondria and superoxide generating NADPH oxidase in breast and ovarian tumors. *Cancer Biol Ther* **4**, 1367–1373.
10. Singh, K. K., Kulawiec, M., Still, I., Desouki, M. M., Geradts, J., and Matsui, S. (2005) Inter-genomic cross talk between mitochondria and the nucleus plays an important role in tumorigenesis. *Gene* **354**, 140–146.
11. Singh, K. K. (2004) Mitochondrial damage checkpoint in apoptosis and genome stability. *FEMS Yeast Res* **2**, 127–132.
12. Rasmussen, A. K., Chatterjee, A., Rasmussen, L. J., and Singh, K. K. (2003) Mitochondria-mediated nuclear mutator phenotype in *Saccharomyces cerevisiae*. *Nucleic Acids Res* **31**, 3909–3917.
13. Delsite, R. L., Rasmussen, L. J., Rasmussen, A. K., Kalen, A., Goswami, P. C., and Singh, K. K. (2003) Mitochondrial impairment is accompanied by impaired oxidative DNA repair in the nucleus. *Mutagenesis* **18**, 497–503.
14. Park, S. Y., Chang, I., Kim, J. Y., Kang, S. W., Park, S. H., Singh, K., and Lee, M. S. (2004) Resistance of mitochondrial DNA depleted cells against cell death: Role of mitochondrial superoxide dismutase. *J Biol Chem* **279**, 7512–7520.
15. Verma, M., Naviaux, R. K., Tanaka, M., Kumar, D., Franceschi, C., and Singh, K. K. (2007) Mitochondrial DNA and cancer epidemiology. *Cancer Res* **67(2)**, 437–9.
16. Booker, L. M., Habermacher, G. M., Jessie, B. C., Sun, Q. C., Baumann, A. K., Amin, M., Lim, S. D., Fernandez-Golarz, C., Lyles, R. H., Brown, M. D., Marshall, F. F., and Petros, J. A. (2006) North American white mitochondrial haplogroups in prostate and renal cancer. *J Urol* **175**, 468–472; discussion 472–463.
17. Canter, J. A., Kallianpur, A. R., Parl, F. F., and Millikan, R. C. (2005) Mitochondrial DNA G10398A polymorphism and invasive breast cancer in African-American women. *Cancer Res* **65**, 8028–8033.
18. Petros, J. A., Baumann, A. K., Ruiz-Pesini, E., Amin, M. B., Sun, C. Q., Hall, J, Lim, S., Issa, M. M., Flanders, W. D., Hosseini, S. H., Marshall, F. F., and Wallace, D. C. (2005) mtDNA mutations increase tumorigenicity in prostate cancer. *Proc Natl Acad Sci USA* **102**, 719–724.
19. Darvishi, K., Sharma, S., Bhat, A. K., Rai, E., and Bamezai, R. N. (2007) Mitochondrial DNA G10398A polymorphism imparts maternal Haplogroup N a risk for breast and esophageal cancer. *Cancer Lett* **249(2)**, 249–55.
20. Bai, R. K., Leal, S. M., Covarrubias, D., Liu, A., and Wong, L. J. (2007) Mitochondrial genetic background modifies breast cancer risk. *Cancer Res* **67(10)**, 4687–94.
21. Wang, L., Bamlet, W. R., de Andrade, M., Boardman, L. A., Cunningham, J. M., Thibodeau, S. N., and Petersen, G. M. (2007) Mitochondrial genetic polymorphisms and pancreatic cancer risk. Cancer Epidemiol. *Biomarkers Prev.* **16(7)**, 1455–9.
22. Fuku, N., Park, K. S., Yamada, Y., Nishigaki, Y., Cho, Y. M., Matsuo, H., Segawa, T., Watanabe, S., Kato, K., Yokoi, K., Nozawa, Y., Lee, H. K., and Tanaka, M. (2007) Mitochondrial haplogroup N9a confers resistance against type 2 diabetes in Asians. *Am J Hum Genet* **80(3)**, 407–15.
23. Alexe, G., Fuku, N., Bilal, E., Ueno, H., Nishigaki, Y., Fujita, Y., Ito, M., Arai, Y., Hirose, N., Bhanot, G., and Tanaka, M. (2007). Enrichment of longevity phenotype in mtDNA haplogroups D4b2b, D4a, and D5 in the Japanese population. *Hum Genet* **121(3–4)**, 347–56.

24. van der Walt, J.M., Nicodemus, K. K., Martin, E. R., Scott, W. K., Nance, M. A., Watts, R. L., Hubble, J. P., Haines, J. L., Koller, W. C., Lyons, K., Pahwa, R., Stern, M. B., Colcher, A., Hiner, B. C., Jankovic, J., Ondo, W. G., Allen, F. H., Jr., Goetz, C. G., Small, G. W., Mastaglia, F., Stajich, J. M., McLaurin, A.C., Middleton, L. T., Scott, B. L., Schmechel, D. E., Pericak-Vance, M. A., and Vance, J. M. (2003) Mitochondrial polymorphisms significantly reduce the risk of Parkinson disease. *Am J Hum Genet* **72(4)**, 804–811.
25. Lee, C. F., Liu, C. Y., Chen, S. M., Sikorska, M., Lin, C. Y., Chen, T. L., and Wei, Y. H. (2005) Mitochondrial genome instability and mtDNA depletion in human cancers. *Ann NY Acad Sci* **1042**, 109–122.
26. Clayton, D. A., and Vinograd, J. (1967) Circular dimer and catenate forms of mitochondrial DNA in human leukaemic leucocytes. *J Pers* **35**, 652–657.
27. Welter, C., Kovacs, G., Seitz, G., and Blin N. (1989) Alteration of mitochondrial DNA in human oncocytomas Genes Chromosome *Cancer* **1**, 79–82.
28. Selvanayagam, P., and Rajaraman, S. (1996) Detection of mitochondrial genome depletion by a novel cDNA in renal cell carcinoma. *Lab Invest* **74**, 592–599.
29. Sangkhathat, S., Chiengkriwate, P., Kusafuka, T., Patrapinyokul, S., and Fukuzawa, M. (2005) Renal cell carcinoma in a pediatric patient with an inherited mitochondrial mutation. *Pediatr Surg Int* **21**, 745–748.
30. Jiang, W. W., Masayesva, B., Zahurak, M., Carvalho, A. L., Rosenbaum, E., Mambo, E., Zhou, S., Minhas, K., Benoit, N., Westra, W. H., Alberg, A., Sidransky, D., Koch, W., and Califano, J. (2005) Increased mitochondrial DNA content in saliva associated with head and neck cancer. *Clin Cancer Res* **11**, 2486–2491.
31. Kim, M. M., Clinger, J. D., Masayesva, B. G., Ha, P. K., Zahurak, M. L., Westra, W. H., and Califano, J. A. (2004) Mitochondrial DNA quantity increases with histopathologic grade in premalignant and malignant head and neck lesions. *Clin Cancer Res* **10**, 8512–8515.
32. Mambo, E., Chatterjee, A., Xing, M., Tallini, G., Haugen, B. R., Yeung, S. C., Sukumar, S., and Sidransky, D. (2005) Tumor-specific changes in mtDNA content in human cancer. *Int J Cancer* **116**, 920–924.
33. Wu, C. W., Yin, P. H., Hung, W. Y., Li, A. F., Li, S. H., Chi, C. W., Wei, Y. H. and Lee, C. (2005) Mitochondrial DNA mutations and mitochondrial DNA depletion in gastric cancer. Genes Chromosomes Cancer 44 , 19–28.
34. Wang, Y., Liu, V. W., Xue, W. C., Tsang, P. C., Cheung, A. N., and Ngan, H. Y. (2005) The increase of mitochondrial DNA content in endometrial adenocarcinoma cells: a quantitative study using laser-captured microdissected tissues. *Gynecol Oncol* **98**, 104–110.
35. Tseng, L. M., Yin, P. H., Chi, C. W., Hsu, C. Y., Wu, C. W., Lee, L. M., Wei, Y. H., and Lee, H. C. (2006) Mitochondrial DNA mutations and mitochondrial DNA depletion in breast cancer. *Genes Chromosomes Cancer* **45**, 629–638.
36. Kulawiec, M., Arnouk, H., Desouki, M. M., Kazim, L., Still, I., Singh, K. K. (2006) Proteomic anlysis of proteins involved in mitochondria-to-nucleus retrograde response in human cancer cells. *Cancer Biol Ther* **8,** 967–975.
37. Vignais, P. V. (2002) The superoxide-generating NADPH oxidase: structural aspects and activation mechanism. *Cell Mol Life Sci* **59**, 1428–1459.
38. Lambeth, J. D. (2004) Oxygen. *Nat Rev Immunol* **4**, 181–189.
39. Cheng, G., Cao, Z., Xu, X., van Meir, E. G., and Lambeth, J. D. (2001) Homologs of gp91phox: cloning and tissue expression of Nox3, Nox4, and Nox5. *Gene* **269**, 131–140.
40. Shidara, Y., Yamagata, K., Kanamori, T., Nakano, K., Kwong, J. Q., Manfredi, G., Oda, H., and Ohta, S. (2005) Positive contribution of pathogenic mutations in the mitochondrial genome to the promotion of cancer by prevention from apoptosis. *Cancer Res* **65**, 1655–1663.
41. Moreno-Loshuertos, R., Acin-Perez, R., Fernandez-Silva, P., Movilla, N., Perez-Martos, A., Rodriguez de Cordoba, S., Gallardo, M. E., and Enriquez, J. A. (2007) Differences in reactive oxygen species production explain the phenotypes *Nat Genet* **38**, 1261–1268.

Chapter 16

Polymorphisms of DNA Repair Genes: ADPRT, XRCC1, and XPD and Cancer Risk in Genetic Epidemiology

Jun Jiang, Xiuqing Zhang, Huanming Yang, and Wendy Wang

Summary

Many studies have suggested that adenosine diphosphate ribosyl transferase (ADPRT), X-ray repair cross-complementing 1 (XRCC1), and xeroderma pigmentosum complementary group D (XPD) are three major DNA base excision repair (BER) genes and that they act interactively in stimulating and executing BER processes. Polymorphisms of these genes may influence the rate of gene transcription, the stability of the messenger RNA, or the quantity and activity of the resulting protein. Thus, the susceptibility or severity of several disorders is influenced by possession of specific alleles of polymorphic genes. So, it is plausible that variations and mutations in these genes affect DNA repair capacity in normal populations, and thus facilitate cancer development in normal or exposed individuals. To promote translation of scientific findings for potential clinical application of DNA repair function, we have searched publications relevant to molecular epidemiology studies of associations between single-nucleotide polymorphisms (SNPs) in the genes, and several frequent human cancer. We have focused on five particular polymorphisms as our starting point: the T→C polymorphism (Val762Ala) in exon 17 of ADPRT, the novel transition at the promoter region (–77T→C) of XRCC1, two common nonsynonymous polymorphisms (Arg194Trp and Arg399Gln), and the C→A silent polymorphism (Arg156Arg) in exon 6 of XPD. We review here the case-control studies examining whether these polymorphisms are correlated with reduced DNA repair efficiency, their influence on the development of different solid tumors, and their possible interactions with other genetic factors and environmental exposures.

Key words: DNA base excision repair, cancer, polymorphisms, ADPRT, XRCC1, XPD, genetic epidemiology.

1. Introduction

DNA in most cells is regularly damaged by endogenous and exogenous mutagens. Unrepaired damage can result in apoptosis or it may lead to unregulated cell growth and cancer *(1)*. Alternatively,

Mukesh Verma (ed.), *Methods in Molecular Biology, Cancer Epidemiology, vol. 471*
© 2009 Humana Press, a part of Springer Science + Business Media, Totowa, NJ
Book doi: 10.1007/978-1-59745-416-2

the damage can be repaired at the DNA level, enabling the cell to replicate as planned. Complex pathways involving numerous molecules have evolved to perform such repair. Because of the importance of maintaining genomic integrity in the general and specialized functions of cells as well as in the prevention of carcinogenesis, genes coding for DNA repair molecules have been proposed as candidate cancer-susceptibility genes *(2–4)*.

At least four pathways of DNA repair operate on specific types of damaged DNA, and each pathway involves numerous molecules. DNA base excision repair (BER) operates on small lesions such as oxidized or reduced bases, fragmented or nonbulky adducts, or those produced by methylating agents *(5)*. Molecules involved with the restoration phase of BER include three major genes: adenosine diphosphate ribosyl transferase (ADPRT), X-ray repair cross-complementing 1 (XRCC1), and xeroderma pigmentosum complementary group D (XPD). Sequence variants in DNA repair genes are thought to modulate DNA repair capacity; consequently, they are suggested to be associated with altered cancer risk *(6)*. Reliable knowledge of which sequence variants influence cancer risk may help in identifying persons at high risk of developing cancer and shed light on cancer etiology. Therefore, we have performed an overall human genomic epidemiology review on associations between genes in the base excision repair pathway and cancer risk, focusing on genes encoding three key enzymes in this repair pathway: ADPRT, XRCC1, and XPD. Our hope is that a consolidation and analysis of current epidemiologic results, combined with advances in our understanding of molecular mechanisms, may help elucidate connections between cancer risk and DNA repair.

2. Methods

We identified relevant articles through a PubMed search of literature published between January 2000 and October 2006 by using the keywords "ADPRT or XRCC1 or XPD," "polymorphism," "cancer," and "population" to scan titles and abstracts in databases. The allele frequencies were obtained mainly from case-control studies. Both qualitative and quantitative studies were considered. Epidemiologic analyses were compiled and reviewed, and polymorphisms in DNA repair genes examined in epidemiologic studies of cancer risk were created (**Table 16.1**). Cancers studied included lung, breast, esophagus, gastric, prostate, cervical, uterine leiomyoma, endometrial, colorectal, pancreas, follicular lymphoma nasopharyngeal, basal cell, oral, and head and neck cancer. For each publication, we extracted information on the

Table 16.1
Polymorphisms in DNA repair genes examined in epidemiologic studies of cancer risk

Gene	Gene ID	MIM	Location	Exon	mRNA position	Mutation	Amino acid	SNP ID
ADPRT	142	173870	1q41-q42	17	2446	T→C	Val762Ala	rs1136410
				—	5′UTR	–77T→C	—	rs3213235
XRCC1	7515	194360	19q13.2	6	685	C→T	Arg194Trp	rs1799782
				10	1301	G→A	Arg399Gln	rs25487
XPD	2068	126340	19q13.3	6	499	C→A	Arg156Arg	rs238406

specimen used, first author, publication year, cancer type, control source, country, ethnic composition of the population, numbers of cases and controls, frequency of allele and genotype, or value of cases for the reported locus, and interaction studied with other genetic and environment. Characteristics of individual studies are summarized in **Tables 16.2–16.6**.

3. Results and Discussion

3.1. Studies of Val762Ala Polymorphism and Cancer Risks in Different Populations

ADPRT is a DNA-binding protein involved in the regulation of BER by detecting DNA strand breaks and poly(ADP-ribosyl)ating, a posttranslational modification of proteins, which relate to DNA repair programs, apoptosis decisions, or both *(6,7)*. Upon binding to DNA, ADPRT becomes auto-poly(ADP-ribosyl)ated, allowing it to interact with other proteins involved in BER. Because the Val762Ala substitution located within the COOH-terminal catalytic domain is associated with deficient poly(ADP-ribosyl)ation activity *(8)*, this amino acid change may reduce the BER capacity, and thereby cause genome instability.

Genetic polymorphisms often vary with ethnicity. The 762Ala allele frequency in Asian populations is about 0.4 *(9,10)*, which is higher than the 0.2 of Caucasians *(11,12)* but lower than the 0.6 of African-Americans *(12)*. Cancer case-control studies show the relationship between Val762Ala polymorphism and cancer risks among that different populations of Chinese and Americans reported (**Table 16.2**). Only Gu et al. *(12)* observed 762Ala allele significantly associated with increased risk (2.9-fold) of lung cancer adenocarcinoma (3.21-fold) in Mexican-Americans

Table 16.2
Val762Ala polymorphisms and various types of cancers in different populations

Specimen used	Cancer type	Ref. No	Country	Cases		Controls		Genotype 1/1for cases (major-allele homozy-gotes)		Genotype 1/2 for cases (heterozy-gotes)		Genotype 2/2 for cases (minor-allele homozygotes)		Interactions studied
				No.	Minor allele frequency	No.	Minor allele frequency	Frequency	OR (95% CI)	fre quency	OR (95% CI)	Frequency	OR (95% CI)	
Buccal cell	Breast	11	American Caucasian	1716	0.168	1371	0.166	0.6958	0.9 (0.6–1.4)	0.2727	1.0 (0.9–1.2)	0.0315	1.0	Undefined
Blood and tissue	Lung	9	China	1,000	0.4385	1,000	0.389	0.307	1.0	0.509	1.17 (0.95–1.43)	0.184	1.68 (1.27–2.23)	The OR of the Ala/Ala genotype for nonsmokers and smokers who smoked <16, 16 to 28, or >28 pack-years were 1.13 (0.79–1.62), 1.35 (0.68–2.70), 2.46 (1.35–4.51) or 17.09 (8.15–35.83). ADPRT 762Ala/Ala + XRCC1 399Gln/Gln: 5.91 (2.09–16.72)

and large cell carcinoma (10.79-fold). Lung cancer attributed to a complex outcome of exposure to endogenous carcinogens, exogenous carcinogens, or both. Oxidative DNA damage also might contribute to the development of lung cancer mainly repaired by the BER pathway. In addition, tobacco smoking contains various carcinogens that can cause DNA damage *(13)*. Theoretically, if the ADPRT 762Ala variant actually diminishes BER capacity, smokers carrying the ADPRT 762Ala/Ala homozygous variant genotype will be at higher risk for lung cancer than those carrying the wild-type genotype. Really, investigators found that a joint effect of the ADPRT 762Ala/Ala genotype and smoking in a dose–response manner. Zhang and colleagues found that the odd ratios (Ors) of the Ala/Ala genotype for nonsmokers and smokers who smoked <16, 16 to 28, or >28 pack-years were 1.13 (0.79–1.62), 1.35 (0.68–2.70), 2.46 (1.35–4.51), and 17.09 (8.15–35.83), respectively *(9,10)*. These studies connect genetic epidemiology and the function of the genetic variations, and they provide biological bases for association between 762Ala/Ala and lung cancer risk.

3.2. Studies of Promoter Region (–77T→C), Arg194Trp, Arg399Gln Polymorphism of XRCC1 Gene and Cancer Risks in Different Populations

The XRCC1 protein plays an important role in BER. After excision of a damaged base, it stimulates endonuclease action and acts as a scaffold in the subsequent restoration of the site *(41)*. Three polymorphisms in *XRCC1* (–77T→C, Arg194Trp, and Arg399Gln) have been examined in epidemiologic studies with fairly consistent results *(9,10,15,18,20,24–26,28–34)*. The distributions of selected characteristics and the XRCC1 –77T→C, Arg194Trp, and Arg399Gln allele frequencies between cancer patients and controls are summarized in **Tables 16.3–16.5**. There are significant differences in the –77T→C polymorphism distributions of genotype between the lung cancer patients and the controls (χ^2 = 14.41, P = 0.0007). Logistic regression analysis revealed that compared with the –77TT wild type, subjects carrying the –77CC homozygote had a 2.98-fold elevated risk with borderline significance (95% confidence interval [CI] = 0.93–9.59). The –77CT/CC variant genotypes were associated with a increased risk for lung cancer (adjusted OR = 1.55; 95% CI = 1.21–1.98) and esophageal squamous cell carcinoma (OR = 1.38; 95% CI = 0.99–1.94).

Investigators examined the association between the XRCC1 –77T→C variant genotypes and risk of cancers by stratifying potential confounding variables, such as age, sex, smoking status and histological types. The risk of lung cancer associated with the XRCC1 –77TC/CC genotypes was more pronounced in subjects with histological types other than in small cell carcinoma, but this difference was not statistically significant (P = 0.6212). The environmental factor smoking seemed to have a relatively effect on risk of lung cancer and esophageal squamous cell carcinoma. Compared with the nonsmokers having the –77TT wild-type

Table 16.3
Promoter region (–77T→C) of XRCC1 gene polymorphisms and various types of cancers in different populations

Specimen used	Cancer type	Ref. No	Country	Cases		Controls		Genotype 1/1 for cases (major-allele homozygotes)		Genotype 1/2 for cases (heterozygotes)		Genotype 2/2 for cases (minor-allele homozygotes)		Interactions studied
				No.	Minor allele frequency	No.	Minor allele frequency	Frequency	OR (95% CI)	Frequency	OR (95% CI)	Frequency	OR (95% CI)	
Blood	Esophagus	10	China	405	0.13	478	0.1	0.753	1.0	0.232	1.38 (0.99–1.94)	0.015	1.46 (0.43–5.03)	Undefined
Blood	Lung	14	China	710	0.1565	710	0.1095	0.704	1.0	0.279	1.51 (1.17–1.94)	0.017	2.98 (0.93–9.59)	9.82-fold increased risk for heavy smokers carrying the –77C variant (–77TC/CC); –77C+C194 + G399 was 1.91 (1.39–2.62); –77T+194T + 399A was 2.02 (1.25–3.26); –77C+194T + 399A was 4.12 (0.46–37.00)

Blood	Breast	54	France	247	0.4185	380	0.401	0.364	1.0	0.433	0.83 (0.57-1.19)	0.202	1.23 (0.71-1.97)	Haplotype (–77C + 194C + 280A + 399G) of XRCC1 was associated with increased breast cancer risk (OR, 1.90; 95% CI, 1.12–3.23)
Blood	Lung	55	China	1024	0.126	1118	0.0925	0.765	1.0	0.218	1.44 (1.16-1.80)	0.017	1.87 (0.87-4.01)	An increased risk of lung cancer was associated with the variant XRCC1 –77 genotypes (TC and CC) compared with the TT genotype (OR =1.46, 95% CI = 1.18–1.82; $P = 0.001$)

OR and 95% CI (Cornfield's method) calculated from published data.

genotype, we observed a 9.82-fold increased risk of lung cancer in heavy smokers (>30 pack-years) carrying the – 77CT/CC genotypes (adjusted OR = 9.82; 95% CI = 5.66–17.02) and a 4.52-fold increased risk (95% CI = 0.51–39.87) for smokers carrying the two variant, suggesting a possible gene–smoking interaction. Moreover, it was noticed that the smokers with the *ADPRT 762 Ala/Ala* and *XRCC1-77 C/C* genotypes had higher risks of esophageal squamous cell carcinoma than nonsmokers. When we combined with the other two common single nucleotide polymorphisms of Arg194Trp on exon 6 and Arg399Gln on exon 10 and performed the haplotypes frequency, the distribution between the cases and the controls was statistically different (χ^2 = 33.16, P < 0.0001). Compared with the most common haplotype TCG, the TTA haplotype was associated with a 2.02-fold increased risk of lung cancer (95% CI = 1.25–3.26) and the CTA haplotype was associated with a 4.12-fold elevated risk (95% CI = 0.46–37.00), which indicated a potential haplotype effect *(10,14)*.

As we now known, –77T→C was identified in the 5′ untranslated region (UTR) of the XRCC1 gene, which may be functional. The same investigators also examined the function of this polymorphism by luciferase gene expression and electrophoretic mobility shift assays, and they found that this –77T→C transition enables or enhances the XRCC1 promoter to bind a potential transcriptional inhibitor, which thus displays a lower promoter activity. If this is the case, the significant association of the XRCC1 –77C allele with an increased risk of lung cancer observed in this molecular epidemiologic study in China may be ascribed to diminished expression of this DNA repair protein due to this newly identified variant.

Table 16.4 shows the most of published studies of Arg194Trp. An increased risk of cancer has been reported to be associated with the Trp allele *(10,14–20)*. The XRCC1 Arg194Trp polymorphism also seemed to be more prevalent among Asians (18.7–43.7% carried the heterozygous variant and 3.25–12.5% carried the homozygous variant) than among Caucasians (8.64–12.7% heterozygous and 0–18.7% homozygous). The frequency of the variant allele was significantly different among the three ethnic groups (Caucasian, 0.187; Asian, 0.326; and African, 0.05; P < 0.05) *(11,18,19)*, which was very similar with the number from a recently published meta-analysis of XRCC1 polymorphisms and cancer risk *(42)*. However, it is unclear that the observed differences by race have a biological basis, and genetic effects for complex diseases tend to be consistent across human populations *(43)*. Therefore, further evidence is needed to make a conclusion about potential differences in relative risk by race.

The largest study for Arg194Trp polymorphism of XRCC1 was a breast cancer research of Americans Caucasians (n = 1580 cases, n = 1253 controls) that showed age and study site-adjusted

Table 16.4
Arg194Trp polymorphisms and various types of cancers in different populations

Specimen used	Cancer type	Ref. no	Country	Cases		Controls		Genotype 1/1 for cases (major-allele homozygotes)		Genotype 1/2 for cases (heterozygotes)		Genotype 2/2 for cases (minor-allele homozygotes)		Interactions studied
				No.	Minor allele frequency	No.	Minor allele frequency	Frequency	OR (95% CI)	Frequency	OR (95% CI)	Frequency	OR (95% CI)	
Blood	Head and neck	15	Korean	120	0.2915	145	0.1685	0.492	1.0	0.433	2.46 (1.41–4.29)	0.075	2.21 (1.34–3.49)	R194W(C-T), R280H(G-A), R399G(G-A) TGG: cancer 0.189, control 0.117 ($P = 0.009$)
Blood	Esophagus	10	china	411	0.3	478	0.3	0.504	1.0	0.394	0.84 (0.63–1.12)	0.102	1.22 (0.74–2.01)	Undefined
Buccal cell	Breast	11	America caucasian	1580	0.0643	1253	0.1866	0.8759	1.0	0.1196	0.9 (0.8-1.2)	0.0044	0.5 (0.2-1.4)	Undefined
Blood	Lung	14	china	710	0.309	710	0.305	0.472	1.0	0.438	1.01 (0.81–1.26)	0.09	1.11 (0.75–1.63)	Undefined

(continued)

Table 16.4 (continued)

Specimen used	Cancer type	Ref. no	Country	Cases		Controls		Genotype 1/1 for cases (major-allele homozygotes)		Genotype 1/2 for cases (heterozygotes)		Genotype 2/2 for cases (minor-allele homozygotes)		Interactions studied
				No.	Minor allele frequency	No.	Minor allele frequency	Frequency	OR (95% CI)	Frequency	OR (95% CI)	Frequency	OR (95% CI)	
Blood	Esophagus	18	Egyptian	48	0.12	48	0.05	0.771	1.0	0.229	2.56 0.73–9.40	0	–	Undefined
	Lung	17	American Caucasian	180	0.0611	461	0.0585	0.8778	1.0	0.1222	1.0 (0.5–1.8)	0	–	
			American African	154	0.0454	243	0.0823	0.9221	1.0	0.0649	0.4 (0.2–0.8)	0.013	2.3 (0.2–25)	Undefined
Blood	Lung	16	china	103	0.321	102	0.245	0.466	1.0	0.427	1.55 (0.8-3.02)	0.107	1.65 (0.48-5.64)	194Trp/Trp + XPD751 Lys allele: 8.77 (1.47–52.31)

Blood and tissue	Undefined	–	0.003	0.87 (0.51–1.49)	0.0774	1.0	0.9196	0.0482	405	0.0417	336	Norway	21	Lung	Blood and tissue
Blood	Arg/Arg194+399 Gln/Gln or Arg/Gln: 2.66(1.44–4.92); 194Trp/Trp or 399Arg/Trp + 399Arg/Arg: 2.74(1.28–5.89); 194Trp/Trp or Arg/Trp+ 399Gln/ Gln or Arg/Gln: 3.64 (1.57–8.46)	2.78 (0.82–9.40)	0.0732	1.89 (0.99–3.62)	0.2846	1.0	0.6423	0.126	123	0.2154	123	Indian	20	Breast	Blood
Blood	XPD exon 6 CA/ AA+XRCC1 194TrpCT/ TT+XRCC3 241Met CC: 3.83 (1.42–10.3); XPD exon 6 CA/AA + XRCC1 194TrpCT/TT + XRCC3 241Met CT/ TT: 9.43 (1.98–44.9)	1.97 (0.86–4.51)	0.151	2.26 (1.20–4.28)	0.472	1.0	0.377	0.3255	164	0.387	106	Thailand	19	Oral	Blood

(continued)

Table 16.4 (continued)

Specimen used	Cancer type	Ref. no	Country	Cases		Controls		Genotype 1/1 for cases (major-allele homozygotes)		Genotype 1/2 for cases (heterozygotes)		Genotype 2/2 for cases (minor-allele homozygotes)		Interactions studied
				No.	Minor allele frequency	No.	Minor allele frequency	Frequency	OR (95% CI)	Frequency	OR (95% CI)	Frequency	OR (95% CI)	
Blood	Gastric	22	Brazilian	160	0.06	150	0.07	0.875	–	0.125	–	0	–	194Trp + 399Gln+XRCC3 Codon 241Met: 8.3 (0.91–100)
Blood and tissue	Lung	23	China	75	0.1866	162	0.3055	0.6667	1.0	0.2933	0.519 (0.74–2.36)	0.0004	0.296 (0.40–2.33)	Undefined
Blood	Colorectal	24	Korean	209	0.345	209	0.321	0.43	1.0	0.45	1.32 (0.74–2.36)	0.12	0.96 (0.40–2.33)	194Trp-280Arg-399Arg: 1.05 (0.69–1.60), 194Arg-280His-399Arg: 1.64 (0.83–3.26), 194Arg-280Arg-399Gln: 1.59 (1.01–2.49)

Blood	Blood
Follicular lymphoma	Nasopharyngeal
26	25
Danish and Swedish	China
428	417
0.0525	0.24
601	495
0.0607	0.31
0.8949	0.557
1.0	1.0
0.1051	0.398
1.0 (0.6-1.5)	0.79 (0.60–1.05)
0	0.045
–	0.48 (0.27–0.86)
Undefined	Undefined

OR and 95% CI (Cornfield's method) calculated from published data.

OR of 0.5 (CT/CC versus TT; 95% CI = 0.2–1.4), and evidence of interactions with menopausal status also showed that those who carried the heterozygous or homozygous T genotype, compared with those who carried the homozygous C genotype, had a reduced risk of breast cancer (OR = 0.8, 95% CI = 0.6–1.0 and OR = 0.5, 95% CI = 0.2–1.4, respectively), suggesting homozygous T was a protecting factor *(11)*. Studies of this polymorphism and lung cancer risk have been inconsistent, with some suggesting an increased risk *(14,16,17)*, a reduced risk *(21,23)*, or no association *(14,21,23)*. Studies of esophagus and gastric cancer also suggest decreased risks when individuals with CT or TT genotype are compared with those with CC genotype *(10,22)*.

A third XRCC1 polymorphism (Arg399Gln) also has been well studied; however, the results suggested associations in different populations for different cancers (**Table 16.5**): decreased risk for breast cancer *(11)*, lung cancer *(14,16,17,21,23)*, oral cancer *(19)*, and colorectal cancer *(35)*; increased risk for lung cancer *(9,34)*, esophageal cancer *(10)*, gastric cancer *(27)*, breast cancer *(20,28)*, pancreatic cancer *(29)*, prostate cancer *(30)*, head and neck cancer *(15)*, gynecological cancer *(31,32)*, colorectal cancer *(18,24)*, nasopharyngeal cancer *(25)*, and follicular lymphoma *(26)*.There were inconsistent results for breast cancer *(11,20,28)*, lung cancer *(9,14,16,17,21,23,34)* and colorectal cancer *(18,24,34)*. XRCC1 Arg399Gln was the most common sequence variant among the three XRCC1 polymorphisms studied, and there was no major variation for the minor allele frequency between Asians and American/European Caucasians (0.20–0.36 versus 0.30–0.47), except for the studies of follicular lymphoma *(26)* and uterine leiomyoma *(32)*, but this was higher than that of African (about 0.15).

We found that most studies were no directly association with cancer risk, but 399Gln/Gln genotype was strongly associated with an increased risk for colorectal cancer (OR = 3.98; $P < 0.001$) among Africans *(18)* and breast cancer (OR = 2.69; $P < 0.05$) among Asians *(20)*. No relationship of 399Gln/Gln was observed with lung cancer; however, the 194 Trp/Trp genotype was associated with a borderline increases risk of lung cancer among Chinese populations (OR = 1.65; 95% CI = 0.48–5.64) and Africans (OR = 2.3; 95% CI = 0.2–25) *(16,17)*. This might be due to the genetic trait differences. It was found that the frequencies of XRCC1 codon 194 Trp allele were significantly higher (32.6%) *(19)* and XRCC1 codon 399Gln was much less (~27%) *(9,10,14–16,23,29,32)* in Chinese population than those in Caucasians (18.7 and 36.6%, respectively) *(11)*. These different frequencies of codon 194 and codon 399 may account for their different contributions to lung cancer risk. It is biologically plausible to assume that XRCC1 polymorphism at codon 194 may have functional significance. However, the sizes of several

Table 16.5
Arg399Gln polymorphisms and various types of cancers in different populations

Specimen used	Cancer type	Ref. No	Country	Cases		Controls		Genotype 1/1 for cases (major-allele homozygotes)		Genotype 1/2 for cases (heterozygotes)		Genotype 2/2 for cases (minor-allele homozygotes)		Interactions studied
				No.	Minor allele frequency	No.	Minor allele frequency	Frequency	OR (95% CI)	Frequency	OR (95% CI)	Frequency	OR (95% CI)	
Blood	Esophagus	10	China	411	0.27	479	0.27	0.543	1.0	0.374	0.77 (0.57–1.04)	0.083	1.20 (0.70–2.05)	762 Ala/Ala of ADPRT + 399Arg/Arg of XRCC1: 7.91 (1.63–38.39)
Buccal cell	Breast	11	America	3039	0.3647	2587	0.3659	0.3995	1.0	0.4715	1.1 (0.9–1.2)	0.129	0.9 (0.8–1.1)	Undefined
Blood and tissue	Lung	9	china	1000	0.2835	1000	0.279	0.535	1.0	0.363	0.93 (0.77–1.13)	0.102	1.17 (0.85–1.61)	ADPRT 762Ala/Ala + XRCC1 399Gln/Gln: 5.91(2.09–16.72)
Blood	Lung	14	china	710	0.268	710	0.2805	0.532	1.0	0.4	0.98 (0.78–1.23)	0.068	0.83 (0.54–1.25)	Undefined

(continued)

Table 16.5 (continued)

Specimen used	Cancer type	Ref. No	Country	Cases		Controls		Genotype 1/1 for cases (major-allele homozygotes)		Genotype 1/2 for cases (heterozygotes)		Genotype 2/2 for cases (minor-allele homozygotes)		Interactions studied
				No.	Minor allele frequency	No.	Minor allele frequency	Frequency	OR (95% CI)	Frequency	OR (95% CI)	Frequency	OR (95% CI)	
Blood	Gastric	27	Poland	281	0.345	390	0.34	0.44	1.0	0.43	0.8 (0.6–1.2)	0.13	1.0 (0.6–1.7)	399GlnGln/Arg-Gln, smoking ≥ 40 pack-year: 4.8 (2.4–4.5)
Blood	Breast	28	China	1088	0.281	1182	0.273	0.516	1.00 (ref)	0.406	0.90 (0.76–1.08)	0.078	1.20 (0.85–1.69)	399Gln/Gln with high level of SHBG: 3.27 (1.16–9.20)
Blood	Pancreatic	29	American	250	0.35	832	0.32	0.44	1.0 (ref)	0.42	1.1 (0.83–1.5)	0.14	1.3 (0.84–2.0)	Arg/Gln + Gln/Gln smoking ≥40 year: women 4.3 (1.6–11.7), men 2.4 (1.1–5.3)

| | | |

Table 16.5 (continued)

Specimen used	Cancer type	Ref. No	Country	Cases		Controls		Genotype 1/1 for cases (major-allele homozygotes)		Genotype 1/2 for cases (heterozygotes)		Genotype 2/2 for cases (minor-allele homozygotes)		Interactions studied
				No.	Minor allele frequency	No.	Minor allele frequency	Frequency	OR (95% CI)	Frequency	OR (95% CI)	Frequency	OR (95% CI)	
Blood	Colorectal	18	Egyptian	48	0.32	48	0.14	0.458	1.0	0.438	3.92 1.40-11.20	0.104	4.20 0.63-34.90	Urban residence with Arg/Gln + Gln/Gln: 9.97 (1.98–43.76)
			American caucasian	180	0.3056	461	0.3611	0.48333	1.0	0.4222	0.8 (0.5–1.2)	0.0944	0.6 (0.3–1.3)	
			American African	154	0.1688	243	0.191	0.6818	1.0	0.2987	1.1 (0.7–1.9)	0.0195	0.6 (0.2–2.3)	Undefined
Blood	Uterine leiomyom	32	Korean female	327	0.272	197	0.0735	0.462	1.0	0.532	6.97 (4.20–10.99)	0.006	2.97 (0.26–34.35)	Undefined

| | | |

Table 16.5 (continued)

Specimen used	Cancer type	Ref. No	Country	Cases		Controls		Genotype 1/1 for cases (major-allele homozygotes)		Genotype 1/2 for cases (heterozygotes)		Genotype 2/2 for cases (minor-allele homozygotes)		Interactions studied
				No.	Minor allele frequency	No.	Minor allele frequency	Frequency	OR (95% CI)	Frequency	OR (95% CI)	Frequency	OR (95% CI)	
Blood and tissue	Non-small cell lung	23	China	75	0.26	162	0.2561	0.5333	1.0	0.4133	1.143 0.646–2.023	0.0533	0.818 0.246–2.726	Undefined
Blood	Gastric	22	Brazilian	160	0.33	150	0.34	0.456	–	0.419	–	0.125	–	194Trp + 399Gln + *XRCC3* Codon 241Met: 8.3 (0.91–100)
Blood and tissue	Non-small cell lung	21	Norway	331	0.352	391	0.3759	0.3897	1.0	0.5166	1.06 (0.77–1.46)	0.0937	0.67 (0.40–1.10)	Undefined

Blood	Follicular lymphoma	26	Danish and Swedish	597	0.0376	417	0.0383	0.9263	1.0	0.0721	0.9 (0.5-1.6)	0.0017	1.2 (0.1-23)	Undefined	
Blood	Nasopharyngeal	25	China	425	0.25	501	0.26	0.567	1.0	0.358	0.82 (0.62–1.08)	0.075	1.20 (0.69–2.06)	Undefined	
Blood	Colorectal	24	Korean	209	0.2535	209	0.196	0.536	1.0	0.421	2.18 (1.23–3.88)	0.043	1.03 (0.31–3.67)	194Trp-280Arg-399Arg: 1.05 (0.69–1.60), 194Arg-280His-399Arg: 1.64 (0.83–3.26), 194Arg-280Arg-399Gln: 1.59 (1.01–2.49)	
Blood	Colorectal	34	China	718	0.252	729	0.2735	0.567	1.0	0.362	0.84 (0.67–1.07)	0.071	0.82 (0.53–1.25)		399Gln + XRCC3 241Thr /Thr: 1.79 (0.99–3.24) 399 Arg/Arg + XRCC3 241Met: 1.95 (1.08–3.52) 399Arg/Arg + 241Thr/Thr: 2.43 (1.21–4.90)

OR and 95% CI (Cornfield's method) calculated from published data.

studies were marginally adequate, and the mechanism of XRCC1 codon 194 polymorphism to lung cancer still needs to be further explored in a larger different populations.

After stratification for variables, such as age, gender, menopausal status, smoking amount, and histologic type of cases, the XRCC1 399Gln genotype was found a significantly increased risk of cancer development associated with smoking ≥40 pack-year *(27,29)*, but one study was excluded *(31)*. This might be explained by the fact that at high levels of exposure, the DNA repair capacity may be saturated, even in individuals having higher repair capacity *(43,44)*. It is possible that such a finding is attributable to chance because of the relatively small numbers in the subgroups. Additional studies with more patients will be needed to confirm this finding. Homozygosity for the variant Gln allele was associated with an elevated risk of postmenopausal breast cancer among subjects with a higher blood level of sex hormone-binding globulin (SHBG) (OR = 3.27; 95% CI = 1.16–9.20) and a reduced risk among those with a lower level of SHBG (OR = 0.60; 95% CI = 0.18–1.97) *(28)*, and urban residence with Gln allele even had a 9.97-fold elevation risk of colorectal caner *(18)*. Most interesting thing is among all of the participants, age and race-adjusted OR for XRCC1 399 and pancreatic cancer were higher among women (OR = 1.3, 95% CI = 0.87–2.1 heterozygotes compared with wild types; OR = 1.6, 95% CI = 0.86–2.9 homozygous variants compared with wild types) compared with men (OR = 0.96, 95% CI = 0.66–1.4 heterozygotes compared with wild types; OR = 1.0, 95% CI = 0.58–1.9 homozygous variants compared with wild types). Observational evidence suggests that estrogens may increase levels of triglycerides and total lipids in human pancreas *(45,46)*, and thus increase the level of reactive oxygen species and subsequent DNA damage in the pancreas. Moreover, endogenous estrogens themselves may damage DNA *(47)* or synergistically increase the DNA-damaging properties of tobacco smoke. Other potential biological explanations include lower DNA repair capacity in women *(48)*.

Because XRCC1 functionally interacts with ADPRT in BER processes, the joint effect between the XRCC1 and ADPRT genotypes in the risk of cancers has been evaluated. A significant interaction between XRCC1 Arg399Gln and ADPRT Val762Ala polymorphisms was observed, with a higher risk of lung cancer (OR = 5.91; 95% CI =2.09–16.72) and esophageal cancer (OR = 7.91; 95% CI = 1.63–38.39) for individuals carrying genotype of ADPRT 762Ala/Ala and XRCC1 399Gln/Gln compared with the combined major alleles. These results clearly indicate that a more than multiplicative joint effect between the ADPRT 762 Ala/Ala and XRCC1 399 Gln/Gln genotype exists in the risk of developing lung cancer *(9,10)*. In head and neck cancer study *(15)*, the frequency of 194Trp-280Arg-399Arg was 0.189

in cases compared with controls of 0.117 ($P = 0.009$); a similar conclusion were reached for colorectal cancer *(24,34)*. And also, the 194Trp allele interaction with 399 Gln associated with the elevation risk of breast cancer and gastric cancer *(20,22)*. All these results suggested that haplotype analysis of multipolymorphism sites would be informative regarding valid content for the association study of common tumors.

3.3. Studies of Arg156Arg of XPD Silent Polymorphism and Cancer Risks in Different Populations

The functional significance of these XPD variants has not yet been elucidated, but some of the variants may be associated with a reduced repair capacity and increased cancer susceptibility. Most of the epidemiologic studies on cancer reported to date have focused on SNPs in codons 312 and 751, because of their high frequencies. But, here we only summarize Arg156Arg (C→A) silent polymorphism. In the whole study group, the distribution of minor alleles do not have significant variance in Asian and American populations (**Table 16.6**), and also this polymorphism was not directly associated with risk of lung cancer *(35,38)*. Because *XPD* polymorphisms are presumed to affect smoking-related cancers by repairing DNA damage caused by tobacco carcinogens, the potential modifying effect of *XPD* genotype on the relationship between these cancers and tobacco smoking is of particular interest. Stratified analyses revealed a significant protective effect of the C allele against lung cancer in never-smokers in one study, the variant allele was associated with increased risk of adenocarcinoma of lung (AA/AC versus CC; OR = 1.65; $P = 0.02$). Furthermore, the presence of one or two variant alleles was associated with increased risk for lung cancer (OR = 2.49; $P = 0.03$) and adenocarcinoma of lung (OR = 5.60; $P = 0.005$) among never-smokers only *(35)*. Another study also observed that the variant genotype was associated with borderline significant increase in risk, especially among the nonsmokers (OR = 4.10; 95% CI = 1.20–14.0; $P < 0.05$) and females (OR = 3.93; 95% CI =1.14–13.6, $P < 0.05$) *(39)*. If DNA repair capacity were important for cancer development, one would expect that the genotype be most important at high exposures for DNA-damaging agents, for example, among heavy smokers. That the effects of the genotype are primarily observed among nonsmokers could indicate that it is other aspects of XPD function that are important or that the biologically effective mutation is in a neighboring gene. Additionally, a large basal cell cancer study suggested an increased risk (AA versus CC/CA, OR = 1.59; 95% CI = 1.02–2.50) that was even higher among individuals without a family history of basal cell cancer *(37)*. The variant genotype was reported to have lower DNA repair capacity than the wild type, although the XPD Arg156Arg (A22541C) polymorphism does not result in an amino acid change. It is therefore unlikely that the enzymatic function of XPD is affected by the mutation.

Table 16.6
The C→A silent polymorphism (Arg156Arg) in exon 6 of XPD polymorphisms and various types of cancers in different populations

Specimen used	Cancer type	Ref. No	Country	Cases		Controls		Genotype 1/1 for cases (major-allele homozygotes)		Genotype 1/2 for cases (heterozygotes)		Genotype 2/2 for cases (minor-allele homozygotes)		Interactions studied
				No.	Minor allele frequency	No.	Minor allele frequency	Frequency	OR (95% CI)	Frequency	OR (95% CI)	Frequency	OR (95% CI)	
Blood	Endometrial	36	American	371	0.4315	420	0.4275	0.315	1.0	0.507	1.06 (0.77-1.46)	0.178	1.03 (0.68-1.56)	312N+K751+Arg156 of XPD: 1.48 (0.88-2.50); D312+K751+Arg156: 1.06 (0.79-1.43)
Blood	Lung	35	china	149	0.49	137	0.478	0.255	1.0	0.5168	1.12 (0.63–1.97)	0.2282	1.04 (0.75–1.44)	Adenocarcinoma of lung with A-allele (AA/AC versus CC; OR = 1.65; P = 0.02). one or two variant A-alleles was associated with increased risk for lung cancer (OR=2.49;P=0.03) and adenocarcinoma of lung (OR=5.60;P=0.005) among nevermokers only.

Blood	Sputum	Cryo-preserved lym-phocytes
Oral	Lung	Basal cell
39	38	37
Thailand	China	Danish
106	119	319
0.3295	0.5378	0.509
164	113	322
0.311	0.4513	0.466
0.425	0.1933	0.2633
1.0	0.47 (0.22–1.01)	1.0
0.491	0.5378	0.4545
1.74 (0.94–3.22)	1.02 0.52–2.01	0.97 (0.67–1.41)
0.085	0.2521	0.2821
0.85 (0.30–2.37)	1.0	1.59 (1.02–2.50)
CA/AA genotype in non-smokers 4.10 (1.20–14.0) and female 3.93 (1.14–13.6); XPD exon 6 CA/AA+XRCC1 194Trp CT/TT+XRCC3 241MetCC: 3.83 (1.42–10.3) ; XPD exon 6 CA/AA+XRCC1 194Trp CT/TT XRCC3 241MetCT/TT: 9.43 (1.98–44.9)	Undefined	Undefined

OR and 95% CI (Cornfield's method) calculated from published data.

The mutation may influence the rate of translation by altering codon usage or reduce XPD protein levels through an effect on mRNA stability *(49,50)*.

In a study of white and black women from America, it was found that the risk A-allele of XPD Arg156Arg is tightly linked to the protective alleles of XPD Asp312Asn and Lys751Gln which increased 1.48-fold the risk of endometrial cancer *(36)*. Three additional studies also observed this relationship *(51–53)*. A linkage analysis of from Thailand also suggested the haploytpes of CA/AA and XRCC1 194 CT/TT combined with XRCC3 241 CC or XRCC3 241 CT/TT increased the 3.83 (1.42–10.3) or 9.43 (1.98–44.9) risk of oral cancer *(39)*. A possibility is that the susceptibility genotype is in linkage disequilibrium with other genotypes that compensate for the conflicting results of single SNP site analysis. Therefore, the combination of multiple sequence variants in the same gene and in genes functioning in the same biochemical pathway might be more important in carcinogenesis.

Appreciable variability in study results needs to be addressed. First, similar to all other case-control studies, inherent biases may lead to spurious findings. Because our study was a hospital-based study, with cases recruited from hospitals and controls recruited from the community population, the study subjects may not be representative of the general population. However, we believe that our results are unlikely to be attributable to selection bias because we used a relatively large number of incident cases and matched the controls to the cases on age, sex, and residential area. Second, although we observed a significant main effect of the variant genotypes, the moderate sample size might have restricted us in identifying significant gene–environment interactions. Finally, the lack of cohesiveness in the results could be due in part to methodological deficiencies in some studies. Therefore, data currently available on BER gene polymorphisms and cancers should be interpreted with caution, and well-designed studies, with adequate statistical power, are warranted to clarify this issue in the future.

The present study shows that five polymorphisms in DNA repair genes are involved in different tumor susceptibility in different populations: the T→C polymorphism (Val762Ala) in exon 17 of ADPRT, the novel transition at the promoter region (–77T→C) of XRCC1, two common nonsynonymous polymorphisms (Arg194Trp and Arg399Gln), and the C→A silent polymorphism (Arg156Arg) in exon 6 of XPD. The interesting point is that enzymes from each of the three genes are known to participate in different DNA repair pathways. This observation is consistent with the induction of multiple DNA lesions by cigarette smoke and by other environmental agents that require different pathways for repair. Therefore, our study represents an important addition to previously published work on polymor-

phism of DNA repair genes and susceptibility of cancer. Further studies with a larger number of subjects and simultaneous measurement of different polymorphic genes are needed to provide a better understanding of the susceptibility phenomenon *(54)*. Additional studies are also needed to clarify the functional significance of BER gene variants.

Acknowledgments

We thank Dr. Mukesh Verma to review and discuss the manuscript.

References

1. Vispe S, Yung TM, Ritchot J, Serizawa H, Satoh MS. (2000) A cellular defense pathway regulating transcription through poly(ADP-ribosyl)ation in response to DNA damage. Proc Natl Acad Sci USA **97(18),** 9886–91.
2. Cairns J. (1982) Aging and cancer as genetic phenomena. Natl Cancer Inst Monogr **60,** 237–9.
3. Knudson AG Jr. (1989) The genetic predisposition to cancer. Birth Defects Orig Artic Ser **25(2),** 15–27. Review
4. Shields PG, Harris CC. (1991) Molecular epidemiology and the genetics of environmental cancer. JAMA **266(5),** 681–7. Review.
5. Goode EL, Ulrich CM, Potter JD. (2002) Polymorphisms in DNA repair genes and associations with cancer risk. Cancer Epidemiol Biomarkers Prev **11(12),** 1513–30. Review. Erratum in: Cancer Epidemiol Biomarkers Prev. 2003 Oct;**12(10),** 1119.
6. Wood RD, Mitchell M, Sgouros J, Lindahl T. Human DNA repair genes. (2001) Science **291(5507),** 1284–9.
7. Yu SW, Wang H, Poitras MF, Coombs C, Bowers WJ, Federoff HJ, Poirier GG, Dawson TM, Dawson VL. (2002) Mediation of poly(ADP-ribose) polymerase-1-dependent cell death by apoptosis-inducing factor. Science **297(5579),** 259–63.
8. Lockett KL, Hall MC, Xu J, Zheng SL, Berwick M, Chuang SC, Clark PE, Cramer SD, Lohman K, Hu JJ. (2004) The ADPRT V762A genetic variant contributes to prostate cancer susceptibility and deficient enzyme function. Cancer Res **64(17),** 6344–8.
9. Zhang X, Miao X, Liang G, Hao B, Wang Y, Tan W, Li Y, Guo Y, He F, Wei Q, Lin D. (2005) Polymorphisms in DNA base excision repair genes ADPRT and XRCC1 and risk of lung cancer. Cancer Res **65(3),** 722–6.
10. Hao B, Wang H, Zhou K, Li Y, Chen X, Zhou G, Zhu Y, Miao X, Tan W, Wei Q, Lin D, He F. (2004) Identification of genetic variants in base excision repair pathway and their associations with risk of esophageal squamous cell carcinoma. Cancer Res. **64(12),** 4378–84.
11. Zhang Y, Newcomb PA, Egan KM, Titus-Ernstoff L, Chanock S, Welch R, Brinton LA, Lissowska J, Bardin-Mikolajczak A, Peplonska B, Szeszenia-Dabrowska N, Zatonski W, Garcia-Closas M. (2006) Genetic polymorphisms in base-excision repair pathway genes and risk of breast cancer. Cancer Epidemiol Biomarkers Prev **15(2),** 353–8.
12. Gu J, Spitz MR, Yang F, Wu X. (1999) Ethnic differences in poly(ADP-ribose) polymerase pseudogene genotype distribution and association with lung cancer risk. Carcinogenesis **20(8),** 1465–9.
13. Hecht SS. (1999) Tobacco smoke carcinogens and lung cancer. J Natl Cancer Inst **91,** 1194–210.
14. Hu Z, Ma H, Lu D, Zhou J, Chen Y, Xu L, Zhu J, Huo X, Qian J, Wei Q, Shen H. (2005)

A promoter polymorphism (–77T>C) of DNA repair gene XRCC1 is associated with risk of lung cancer in relation to tobacco smoking. Pharmacogenet Genomics **15(7),** 457–63.

15. Tae K, Lee HS, Park BJ, Park CW, Kim KR, Cho HY, Kim LH, Park BL, Shin HD. (2004) Association of DNA repair gene XRCC1 polymorphisms with head and neck cancer in Korean population. Int J Cancer **111(5),** 805–8.
16. Chen S, Tang D, Xue K, Xu L, Ma G, Hsu Y, Cho SS. (2002) DNA repair gene XRCC1 and XPD polymorphisms and risk of lung cancer in a Chinese population. Carcinogenesis **23(8),** 1321–5.
17. David-Beabes GL, London SJ. (2001) Genetic polymorphism of XRCC1 and lung cancer risk among African-Americans and Caucasians. Lung Cancer. **34(3),** 333–9.
18. Abdel-Rahman SZ, Soliman AS, Bondy ML, Omar S, El-Badawy SA, Khaled HM, Seifeldin IA, Levin B. (2000) Inheritance of the 194Trp and the 399Gln variant alleles of the DNA repair gene XRCC1 are associated with increased risk of early-onset colorectal carcinoma in Egypt. Cancer Lett **159(1),** 79–86.
19. Kietthubthew S, Sriplung H, Au WW, Ishida T. (2006) Polymorphism in DNA repair genes and oral squamous cell carcinoma in Thailand. Int J Hyg Environ Health **209(1),** 21–9.
20. Chacko P, Rajan B, Joseph T, Mathew BS, Pillai MR. (2005) Polymorphisms in DNA repair gene XRCC1 and increased genetic susceptibility to breast cancer. Breast Cancer Res Treat **89(1),** 15–21.
21. Zienolddiny S, Campa D, Lind H, Ryberg D, Skaug V, Stangeland L, Phillips DH, Canzian F, Haugen A. (2006) Polymorphisms of DNA repair genes and risk of non-small cell lung cancer. Carcinogenesis **27(3),** 560–7.
22. Duarte MC, Colombo J, Rossit AR, Caetano A, Borim AA, Wornrath D, Silva AE. (2005) Polymorphisms of DNA repair genes XRCC1 and XRCC3, interaction with environmental exposure and risk of chronic gastritis and gastric cancer. World J Gastroenterol **11(42),** 6593–600.
23. Chan EC, Lam SY, Fu KH, Kwong YL. (2005) Polymorphisms of the GSTM1, GSTP1, MPO, XRCC1, and NQO1 genes in Chinese patients with non-small cell lung cancers: relationship with aberrant promoter methylation of the CDKN2A and RARB genes. Cancer Genet Cytogenet **162(1),** 10–20.
24. Hong YC, Lee KH, Kim WC, Choi SK, Woo ZH, Shin SK, Kim H. (2005) Polymorphisms of XRCC1 gene, alcohol consumption and colorectal cancer. Int J Cancer **116(3),** 428–32.
25. Cao Y, Miao XP, Huang MY, Deng L, Hu LF, Ernberg I, Zeng YX, Lin DX, Shao JY. (2006) Polymorphisms of XRCC1 genes and risk of nasopharyngeal carcinoma in the Cantonese population. BMC Cancer **6,** 167.
26. Smedby KE, Lindgren CM, Hjalgrim H, Humphreys K, Schollkopf C, Chang ET, Roos G, Ryder LP, Falk KI, Palmgren J, Kere J, Melbye M, Glimelius B, Adami HO. (2006) Variation in DNA repair genes ERCC2, XRCC1, and XRCC3 and risk of follicular lymphoma. Cancer Epidemiol Biomarkers Prev **15(2),** 258–65.
27. Huang WY, Chow WH, Rothman N, Lissowska J, Llaca V, Yeager M, Zatonski W, Hayes RB. (2005) Selected DNA repair polymorphisms and gastric cancer in Poland. Carcinogenesis **26(8),** 1354–9.
28. Shu XO, Cai Q, Gao YT, Wen W, Jin F, Zheng W. (2003) A population-based case-control study of the Arg399Gln polymorphism in DNA repair gene XRCC1 and risk of breast cancer. Cancer Epidemiol Biomarkers Prev **12(12),** 1462–7.
29. Duell EJ, Holly EA, Bracci PM, Wiencke JK, Kelsey KT. (2002) A population-based study of the Arg399Gln polymorphism in X-ray repair cross-complementing group 1 (XRCC1) and risk of pancreatic adenocarcinoma. Cancer Res **62(16),** 4630–6.
30. Chen L, Ambrosone CB, Lee J, Sellers TA, Pow-Sang J, Park JY. (2006) Association between polymorphisms in the DNA repair genes XRCC1 and APE1, and the risk of prostate cancer in white and black Americans. J Urol. **175(1),** 108–12; discussion 112.
31. Niwa Y, Matsuo K, Ito H, Hirose K, Tajima K, Nakanishi T, Nawa A, Kuzuya K, Tamakoshi A, Hamajima N. (2005) Association of XRCC1 Arg399Gln and OGG1 Ser-326Cys polymorphisms with the risk of cervical cancer in Japanese subjects. Gynecol Oncol **99(1),** 43–9.
32. Jeon YT, Kim JW, Park NH, Song YS, Kang SB, Lee HP. (2005) DNA repair gene XRCC1 Arg399Gln polymorphism is associated with increased risk of uterine leiomyoma. Hum Reprod **20(6),** 1586–9. Epub 2005 Mar 10.
33. Park JY, Lee SY, Jeon HS, Bae NC, Chae SC, Joo S, Kim CH, Park JH, Kam S, Kim IS, Jung TH. (2002) Polymorphism of the DNA repair gene XRCC1 and risk

of primary lung cancer. Cancer Epidemiol Biomarkers Prev **11(1),** 23–7.

34. Yeh CC, Sung FC, Tang R, Chang-Chieh CR, Hsieh LL. (2005) Polymorphisms of the XRCC1, XRCC3, & XPD genes, and colorectal cancer risk: a case-control study in Taiwan. BMC Cancer **5,** 12.
35. Yin J, Li J, Ma Y, Guo L, Wang H, Vogel U. (2005) The DNA repair gene ERCC2/XPD polymorphism Arg 156Arg (A22541C) and risk of lung cancer in a Chinese population. Cancer Lett **223(2),** 219–26.
36. Ketelslegers HB, Gottschalk RW, Godschalk RW, Knaapen AM, van Schooten FJ, Vlietinck RF, Kleinjans JC, van Delft JH. (2006) Interindividual variations in DNA adduct levels assessed by analysis of multiple genetic polymorphisms in smokers. Cancer Epidemiol Biomarkers Prev **15(4),** 624–9.
37. Vogel U, Olsen A, Wallin H, Overvad K, Tjonneland A, Nexo BA. (2005) Effect of polymorphisms in XPD, RAI, ASE-1 and ERCC1 on the risk of basal cell carcinoma among Caucasians after age 50. Cancer Detect Prev **29(3),** 209–14.
38. Shen M, Berndt SI, Rothman N, Demarini DM, Mumford JL, He X, Bonner MR, Tian L, Yeager M, Welch R, Chanock S, Zheng T, Caporaso N, Lan Q. (2005) Polymorphisms in the DNA nucleotide excision repair genes and lung cancer risk in Xuan Wei, China. Int J Cancer **116(5),** 768–73.
39. Kietthubthew S, Sriplung H, Au WW, Ishida T. (2006) Polymorphism in DNA repair genes and oral squamous cell carcinoma in Thailand. Int J Hyg Environ Health **209(1),** 21–9.
40. Vidal AE, Boiteux S, Hickson ID, Radicella JP. (2001) XRCC1 coordinates the initial and late stages of DNA abasic site repair through protein-protein interactions. EMBO J **20(22),** 6530–9.
41. Hu Z, Ma H, Chen F, Wei Q, Shen H. (2005) XRCC1 polymorphisms and cancer risk: a meta-analysis of 38 case-control studies. Cancer Epidemiol Biomarkers Prev **14,** 1810 – 8.
42. Ioannidis JP, Ntzani EE, Trikalinos TA. (2004) 'Racial' differences in genetic effects for complex diseases. Nat Genet **36(12),** 1312–8.
43. Vineis, P. (1997) Molecular epidemiology: low-dose carcinogens and genetic susceptibility. Int J Cancer **71(1),** 1–3.
44. Nakachi K, Imai K, Hayashi S, Watanabe J, Kawajiri K. (1991) Genetic susceptibility to squamous cell carcinoma of the lung in relation to cigarette smoking dose. Cancer Res **51(19),** 5177–80.
45. Tiscornia OM, Cresta MA, de Lehmann ES, Belardi G, Dreiling DA. (1986) Estrogen effects on exocrine pancreatic secretion in menopausal women: a hypothesis for menopause-induced chronic pancreatitis. Mt Sinai J Med **53(5),** 356–60.
46. Tiscornia OM, Cresta MA, de Lehmann ES, Celener D, Dreiling DA. (1986) Effects of sex and age on pancreatic secretion. Int J Pancreatol **1(2),** 95–118.
47. Burcham PC. (1999) Internal hazards: baseline DNA damage by endogenous products of normal metabolism. Mutat Res **443(1–2),** 11–36.
48. Wei Q, Cheng L, Amos CI, Wang LE, Guo Z, Hong WK, Spitz MR. (2000) Repair of tobacco carcinogen-induced DNA adducts and lung cancer risk: a molecular epidemiologic study. J Natl Cancer Inst. **92(21),** 1764–72.
49. Dybdahl M, Vogel U, Frentz G, Wallin H, Nexo BA. (1999) Polymorphisms in the DNA repair gene XPD: correlations with risk and age at onset of basal cell carcinoma. Cancer Epidemiol Biomarkers Prev **8(1),** 77–81.
50. Vogel U, Hedayati M, Dybdahl M, Grossman L, and Nexo BA. (2001) Polymorphisms of the DNA repair gene XPD: correlations with risk of basal cell carcinoma revisited. Carcinogenesis **22(6),** 899–904. Erratum in: Carcinogenesis 2002 Feb;**23(2),** 373.
51. Vogel U, Laros I, Jacobsen NR, Thomsen BL, Bak H,Olsen A,Bukowy Z,Wallin H, Overvad K, Tjonneland A, Nexo BA, Raaschou-Nielsen O. (2004) Two regions in chromosome 19q13.2–3 are associated with risk of lung cancer. Mutat Res **546 (1–2),** 65–74.
52. Zhou W, Liu G, Miller DP, Thurston SW, Xu LL, Wain JC, Lynch TJ, Su L, Christiani DC. (2002) Gene-environment interaction for the ERCC2 polymorphisms and cumulative cigarette smoking exposure in lung cancer. Cancer Res **62(5),** 1377–81.
53. Hou SM, Falt S, Angelini S, Yang K, Nyberg F, Lambert B, and Hemminki K. (2002) The XPD variant alleles are associated with increased aromatic DNA adduct level and lung cancer risk. Carcinogenesis **23(4),** 599–603.
54. Au WW, Salama SA. (2005) Use of biomarkers to elucidate genetic susceptibility to cancer. Environ Mol Mutagen **45(2–3),** 222–8. Review.

Chapter 17

Risk Factors and Gene Expression in Esophageal Cancer

Xiao-chun Xu

Summary

Esophageal cancer is a significant worldwide health problem because of its poor prognosis and high incidence in certain parts of the world. Tobacco smoke and alcohol consumption are significant risk factors for esophageal squamous cell carcinoma, whereas frequent gastroesophageal reflux and subsequent inflammatory reactions play a role in causing the adenocarcinoma. Esophageal carcinogenesis involves multiple genetic alterations. A large body of knowledge has been generated regarding molecular alterations associated with esophageal carcinogenesis. These alterations include aberrant cell cycle control, DNA repair, cellular enzymes, growth factor receptors, and nuclear receptors. This chapter reviews the most frequent gene alterations and their correlation with risk factors as well as the prevention strategies in esophageal cancer.

Key words: Esophageal cancer, risk factor, tobacco smoke, gastroesophageal reflux, gene alterations, cancer prevention.

1. Introduction

Esophageal cancer is a significant worldwide health problem, with a typically advanced pathologic stage at diagnosis, and it is associated with a very poor prognosis *(1–5)*. Although relatively less common in the United States than in other countries, esophageal cancer accounts for an estimated 15,560 new cases and 13,940 deaths, and it was the sixth leading cause of death from cancers among American men in 2007 *(6)*. The most common histologic types of primary esophageal cancer are squamous cell carcinoma (SCC) and adenocarcinoma, which have relatively different epidemiology and risk factors (see subheading No. 2). The incidence of esophageal SCC shows a greater variation in geographic

Mukesh Verma (ed.), *Methods in Molecular Biology, Cancer Epidemiology, vol. 471*
© 2009 Humana Press, a part of Springer Science + Business Media, Totowa, NJ
Book doi: 10.1007/978-1-59745-416-2

distribution than any other malignancy, and esophageal SCC occurs at very high rates in parts of China, Iran, South Africa, France, and Italy (see review in **ref.** *2)*. For example, Linxian and surrounding counties in China have one of the highest incidence and mortality rates for esophageal SCC in the world. Conversely, epidemiologic studies have shown that the incidence of esophageal adenocarcinoma has increased rapidly in the past three decades in the United States and some other Western countries, for reasons that are not clear *(3–5)*. The factors underlying this changing pattern of the neoplasm are unknown, but Barrett's esophagus may be associated with the increase. This esophageal lesion is considered a premalignant lesion, because patients with this disease have a greatly (30–125 times) increased risk of developing esophageal adenocarcinoma *(7–9)*. Currently, esophageal adenocarcinoma accounts for approximate two thirds of the esophageal cancer cases in the United States. In addition, there are marked differences in esophageal cancer incidence among racial groups in the United States: the incidence of esophageal SCC in black men is 5 times more than that for white men, but esophageal adenocarcinoma occurs more often among white persons *(8,10)*. Also the incidence of esophageal carcinoma among blacks is greater at younger ages *(8)*.

The actual etiology of esophageal cancer remains unclear, but tobacco smoke and alcohol consumption are significant risk factors for esophageal SCC, whereas nutrition factors, such as consumption of pickled vegetables or micronutrition deficiency, may contribute to esophageal SCC in certain parts of the world such as China *(1,2)*. In contrast, esophageal adenocarcinoma is more likely to occur in individuals who have frequent gastroesophageal reflux or gastroesophageal reflux disease (GERD), which usually results in Barrett's esophagus *(3–5,10,11)*. The latter is characterized by the replacement of the squamous cell epithelium in the lower to middle esophagus with columnar epithelium (incomplete intestinal metaplasia with or without dysplastic lesions). Furthermore, a high body mass index (BMI) may also be linked to the increasing incidence of esophageal adenocarcinoma, because BMI is associated with gastroesophageal reflux and the development of Barrett's metaplasia *(12,13)*. In addition, tobacco smoke may place obese smokers into a very high-risk group for developing Barrett's esophagus and adenocarcinoma. Tobacco smoke also can enhance the effects of bile acid (a gastrointestinal tumor promoter present in gastroesophageal reflux) in causing esophageal adenocarcinoma *(13)*.

Development of esophageal cancer, like all other cancers, involves multiple genetic alterations, e.g., oncogene activation and tumor-suppressor gene dysfunction *(1–5,14)*. Over the past several decades, a large body of knowledge has developed regarding molecular alterations associated with esophageal

carcinogenesis. These alterations include aberrant cell cycle control, DNA repair, cellular enzymes, growth factor receptors, and nuclear receptors *(1–5,7,14,15)*. These disparate observations have begun to coalesce into an improved understanding of this disease, with some implications for prevention, early detection, and treatment. For example, studies from our laboratory and others have linked esophageal cancer risk factors and the altered gene expressions in esophageal cancer *(16–19)*. This chapter reviews the most frequent gene alterations and their correlation with risk factors in esophageal cancer as well as the prevention of esophageal cancer.

2. Risk Factors of Esophageal Cancer

Esophageal SCC and adenocarcinoma have relatively different risk factors. For example, dietary and environmental factors have strongly been implicated in the etiology of esophageal SCC *(1,2)*, whereas frequent gastroesophageal reflux and subsequent inflammatory reactions play an important role in causing the adenocarcinoma *(3,4)*. Whether these two types of esophageal cancer differ in their natural history and treatment outcome remains controversial. The etiologic and molecular differences between these two types of esophageal cancer should justify differentiated therapeutic considerations *(20)*, although presently, similar treatment and clinical outcome are reported.

2.1. Tobacco Smoke and Esophageal Carcinogenesis

Many studies of epidemiologic data and laboratory experiments have demonstrated that tobacco smoking is associated with cancer development, including esophageal cancer *(21–24)*. In particular, tobacco smoke contains at least 2,550 known compounds, >60 of which have been found to be carcinogenic for humans and/or experimental animals *(25–27)*. Among these carcinogens, polycyclic aromatic hydrocarbons such as benzo[a]pyrene diol epoxide (BPDE) and *N*-nitrosamines are considered to be the most important in tobacco smoke *(25–27)*. Tobacco smoke is causally associated with both esophageal SCC and adenocarcinoma *(21–24)*. The mechanisms of their actions are believed to induce DNA adducts, gene methylation and mutation, and chromosome translocation in target organs *(25–28)*. For example, benzo[a]pyrene undergoes metabolic biotransformation to electrophilic intermediates such as BPDE that react with cellular macromolecules, forming DNA adducts, which are carcinogen metabolites covalently bound to DNA, usually at guanine or adenine residues *(26–28)*. Major advances in the understanding of DNA adduct structure and its consequences over the past two decades indicate that

if DNA adducts escape cellular repair mechanisms and persist, they may lead to miscoding and result in a permanent mutation *(26,27,29)*. Cells with damaged DNA may be removed by apoptosis ***(30)***. Alternatively, if a permanent mutation occurs in a critical region, an oncogene may be activated or a tumor-suppressor gene deactivated. Multiple events of this type lead to aberrant cells with loss of normal growth control, migration, and adhesion and ultimately to cancer *(26,27,29)*. Several studies have reported direct evidence of tobacco carcinogen-targeted genes, that is, hot spot mutations in k-ras and p53 genes *(16,31)*, which correlated with benzo[a]pyrene exposure. Cigarette smoke also causes morphologic changes and the loss of retinoic acid receptor (RAR)-β_2 expression in lung tissues of animals *(32)*, and cigarette smoke and the tobacco carcinogen 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone have been shown to induce the methylation of the RAR-β_2 gene promoter in murine lung cancer *(33)*. Our studies similarly showed that BPDE, a carcinogen present in tobacco smoke and environmental pollution, suppressed RAR-β_2 expression in premalignant and malignant esophageal cells and that BPDE induced the methylation of the RAR-β_2 gene promoter *(17,19)*. Consequently, expression of epidermal growth factor receptor (EGFR), extracellular signal-regulated kinase (ERK) 1/2, activated protein (AP)-1, and cyclooxygenase (COX)-2 was induced by BPDE *(19)*. However, epigallocatechin gallate, a chemopreventive agent extracted from the tea plant, has been shown to bind to DNA methyltransferases and reactivate RAR-β_2 expression in esophageal cancer cells *(34)*. Our additional study on the direct effects of BPDE on gene expression used our newly developed DNA-immunoprecipitation technique and identified and cloned BPDE-binding DNA fragments. After their sequences were blasted in the GenBank database, these fragments included ataxia telangiectasia mutated (ATM), breast cancer (BRCA) gene-1, apoptosis-related genes, cellular enzymes, expressed sequence tag (EST) clones, and cytosine-phosphoguanosine (CpG) islands *(35*; our unpublished data). The results of these studies demonstrate that tobacco smoke can induce gene alterations and that the latter are important in causing esophageal carcinogenesis and cancer progression.

2.2. Alcohol Consumption and Esophageal SCC

Alcohol is another major risk factor causing esophageal SCC *(22,23)*. Chronic and excessive consumption of alcohol can impair the body's biochemical metabolism and nutrition status and alter gene expression in target tissues *(36)*. In the human body, ethanol is mainly metabolized by alcohol dehydrogenase in the liver, which results in generation of acetaldehyde. The latter is toxic and carcinogenic because acetaldehyde may induce gene mutation and inhibits retinoic acid biosynthesis *(36–38)*. Simultaneous ethanol administration had an enhancing

effect and also a promoting effect in the postinitiation phase on *N*-nitrosomethylbenzylamine (NMBA)-induced rat tumorigenesis *(39)*. Furthermore, microsomal cytochrome P-4502E1 (CYP2E1)-dependent ethanol metabolism can produce reactive oxygen species (ROS), leading to cell injury, inflammation, and lipid peroxidation *(36,40)*. Induction of CYP2E1 enzyme by alcohol can enhance degradation of retinal and retinoic acid *(36,40)*. Reduced retinoic acid levels in the cells result in an altered expression of different genes, such as reduced expression of RAR-β_2, but in an increase in expression of EGFR, Erk1/2, AP-1, and COX-2 *(19,41)*. Alcohol consumption also may enhance the role of tobacco carcinogens in esophageal cancer development, because these two are often used at the same time by patients with esophageal SCC.

2.3. Gastroesophageal Reflux, Barrett's Esophagus, and Esophageal Adenocarcinoma

Chronic gastroesophageal reflux or GERD usually leads to replacement of squamous mucosa (normal lining of the esophagus) with metaplastic, intestinal-type Barrett's mucosa, a condition termed Barrett's esophagus *(3,4,7)*. The extent of intestinal metaplasia in the esophagus is directly related to the severity of the underlying GERD *(42)*. Gastroesophageal reflux itself is not a disease but a normal physiological process. It becomes pathologic (i.e., GERD) if gastroesophageal reflux produces symptoms and tissue injury in the esophagus, oropharynx, larynx, or respiratory tract *(7)*. The etiology of GERD is not clear. Tobacco smoke, alcohol consumption, hot beverages, and certain food (such as pizza) may trigger gastroesophageal reflux *(7)*. Overweight and obesity also causes gastroesophageal reflux or GERD, and treatment of obesity may effectively reduce GERD and esophageal adenocarcinoma *(43–45)*. Obese persons have increased intraab dominal pressure that displaces the lower esophageal sphincter and increases the gastroesophageal gradient, and they also have a higher output of bile and pancreatic enzymes, which makes the refluxate more toxic to the esophageal mucosa *(45)*. In addition, a weakened esophageal defense mechanism also may contribute to tissue injury by the refluxate, which leads to GERD *(7)*.

Barrett's esophagus is an acquired metaplastic change that occurs in the distal esophagus secondary to chronic gastroesophageal reflux, and it is a multiphase process *(7)*. In the initiation phase, squamous cells around the gastroesophageal junction suffer clinical or occult reflux damage and/or tobacco smoke irritation and change to a new cell phenotype (incomplete intestinal metaplasia), and this type of cell is more resistant to acid and bile-caused injuries. Therefore, short- or long-segment Barrett's esophagus is formed in the formation phase *(46)*. This metaplastic epithelium either remains dormant or progresses to dysplasia and adenocarcinoma during the progression phase *(7,46)*. Acid and

bile reflux variably affect Barrett's esophagus, and they may cause dysplasia or adenocarcinoma *(46,47)*. Acid may be synergistic to bile, and the latter is thought to be an endogenous promoter of gastrointestinal cancer because bile acid can activate the mitogen-activated protein kinase (MAPK) pathway and induce AP-1 and COX-2 expression *(46–49)*; these genes have an antiapoptotic role in cells and in turn increase cell proliferation, promote the invasiveness potential of Barrett's metaplasia and neoplasia, or both *(46–49)*. In particular, our data showed that bile acid inhibited RAR-β_2 expression and induced EGFR and AP-1 levels and COX-2 expression in normal and cancerous esophageal cells *(50*; our unpublished observation). Others found that acid and bile acid induced ERK activity, peroxisome proliferator-activated receptor-γ expression, and cell proliferation in normal esophageal epithelial cells *(51)*; that bile acid exposure caused phosphatidylinositol-3-kinase-mediated proliferation in Barrett's adenocarcinoma cells *(52)*, and that deoxycholic acid at neutral pH-activated nuclear factor-κB (NF-κB) and induced interleukin-8 expression in esophageal cells in vitro *(53)*. In addition, other factors, such as proinflammatory cytokines, NF-κB, inducible nitric-oxide synthase (iNOS), and growth factors also play a significant role in the development and progression of esophageal adenocarcinoma *(54,55)*.

2.4. Additional Risk Factors and Esophageal Carcinogenesis

Owing to the high geographic variation in esophageal cancer incidence, it is speculated that some localized risk factors may play an important role in esophageal carcinogenesis. For example, fungal toxins and spices are thought to have a positive correlation with the development of esophageal SCC. Home-brewed beers and other spirits contaminated with mycotoxins cause esophageal SCC in African countries *(2,7)*. Human papilloma virus serotypes 16 and 18 also have been found to be a risk factor for esophageal SCC *(2,7,56,57)*. However, certain types of *Helicobacter pylori* infection have protective efforts on esophageal cancer development, although the infection is associated with an increased incidence of gastric cancer *(58,59)*. In addition, pickled vegetables and micronutrition deficiency (such as zinc) or overload (such as heme iron) also may contribute to esophageal SCC in some parts of China and in experimental animals *(2,24,60,61)*. High tissue zinc concentration was strongly associated with a reduced risk of developing esophageal SCC *(60,61)*, and zinc deficiency in rats enhances esophageal cell proliferation, causes alteration in gene expression, and promotes esophageal carcinogenesis *(62)*. Finally, other pathological conditions (such as achalasia, esophageal webs, or Fanconi anemia) also are associated with increases in esophageal cancer *(7,63,64)*.

3. Molecular Alternations in Esophageal Cancer

Molecular studies of esophageal cancer have revealed frequent genetic abnormalities in esophageal SCC and adenocarcinoma (**1**–*5*,*7*,*65*,*66*), such as altered expression of p53, p16, cyclin D1, EGFR, E-cadherin, COX-2, iNOS, RARs, ATM, Rb, hTERT, p27, p21, APC, c-MYC, vascular endothelial growth factor (VEGF), TGT-α, NF-κB, and others. In many cases, certain gene alterations are consistently observed in both esophageal SCC and adenocarcinoma, but in some cases aberrant expression of genes is more frequently observed in one type of esophageal cancer than in another type. The altered gene expression is often closely correlated with the risk factors in esophageal cancer. In this section, the most altered genes and their role in esophageal cancer development are discussed in more detail.

3.1. p53 Gene Mutation and Esophageal Cancer

p53 is a tumor-suppressor gene, and its primary function is to maintain human genetic stability and DNA repair capacity *(67)*. Tobacco smoke can cause p53 gene mutation, which is also one of the most frequent events in esophageal carcinogenesis *(67)*. The incidences of p53 mutations in both types of esophageal cancer have been reported to be quite high *(15*,*68)*. The p53 protein level is progressively increased from normal via hyperplasia to dysplasia of the esophagus *(67*,*69)*, indicating that p53 plays a role in suppression of esophageal carcinogenesis, although its prognostic value remains controversial. Some studies have noted a worse prognosis when p53 is overexpressed *(70*,*71)*, whereas other studies have noted a shorter survival when p53 is not expressed *(72)*. Other studies have failed to show a survival correlation with p53 *(73*,*74)*. The reason for this discrepancy may be due to the methods used for detection of p53 gene mutations, that is, immunohistochemical detection of p53 protein overexpression usually indicates mutation of the p53 gene because the mutated p53 protein has a longer half-life than that of wild-type p53 protein *(67)*. However, the frame shift and deletion mutations of the p53 gene result in lost expression of p53 protein, which will lead to negative immunohistochemical data. Therefore, detection of p53 protein overexpression by immunohistochemistry is not a very useful method in predicting the p53 gene mutation, and direct DNA sequencing is essential.

3.2. p16 Gene Alteration and Esophageal Cancer

p16 protein is a cell cycle regulator and inhibits cyclin-dependent kinases (CDKs) 4 and 6 and so limits CDK 4/6-cyclin D complex formation and phosphorylation of Rb. This blocks cell cycle progression from the G1 phase to the S phase *(75)*. Alterations in the p16 expression lead to deregulation of cell proliferation and

consequent genomic instability *(75)*. Inactivation of the p16 gene occurs by three different mechanisms: hypermethylation of the gene promoter, mutation coupled with loss of the second allele, and homozygous deletion *(76)*. Mutation, loss of heterozygosity (LOH), or promoter hypermethylation of the p16 gene occurs in up to 80% in patients with premalignant lesions and esophageal cancer *(1–5)*. In particular, p16 hypermethylation correlated with the degree of dysplasia in specialized intestinal metaplasia *(75,77,78)*. p16 gene promoter methylation was commonly detected not only in earlier stages of esophageal SCC but also in precancerous lesions and in the background normal-appearing epithelia *(79)*, that is, immunoexpression of p16 protein was lost in 38 (90.5%) of 42 esophageal SCCs and in 34 (81%) of 42 dysplastic lesions, whereas the histologically normal epithelia adjacent to the tumor showed p16 expression even in the presence of p16 methylation. Further investigation confirmed that p16 silencing was either through methylation or through LOH and possible mutation in these esophageal SCC or dysplastic lesions of the esophagus *(79)*. LOH at the p16 gene locus was found in 68% of esophageal adenocarcinomas and in 55% of the premalignant lesions. Of 28 samples without p16 immunoreactivity, 25 (89%) showed p16 promoter hypermethylation with or without LOH at the p16 gene locus. This study indicated that p16 gene promoter methylation was the predominant mechanism for p16 inactivation *(75)*, although another study found that two thirds of esophagus squamous cancer cell lines have deletions in the p16 gene *(80)*. Moreover, 9p21 LOH, p16 gene promoter methylation, and p16 mutations were detected in 57, 61, and 15%, respectively, of 107 patients with Barrett's esophagus *(81)*. Furthermore, introduction of p16 expression significantly suppressed cell proliferation in the p16-deleted esophageal SCC cell line. VEGF secretion was significantly suppressed in p16-transfected tumor cells compared with nontransfected tumor cells *(81)*. In addition, a study demonstrated that chemoprevention agents such as genistein can reverse DNA hypermethylation and reactivate RAR-β_2, p16, and *O*-6-methylguanine-DNA methyltransferase in esophageal cancer cell lines. Genistein dose-dependently inhibited DNA methyltransferase activity, showing substrate- and methyl donor-dependent inhibition. Genistein also inhibited tumor cell growth *(82)*.

3.3. Altered Cyclin D1 Expression and Esophageal Cancer

Cyclin D1 protein plays a critical role in progression of the cell cycle through the G1 phase. Cyclin D1 overexpression in esophageal SCC has been reported in 23 to 73% of tumor samples *(1–5)*; however, the expression level seems to be relatively constant between dysplastic lesions and carcinoma *(83)*. In addition, cyclin D1 expression correlated with markers of proliferation, suggesting that this protein may be vital for tumor progression *(84)*. Overexpression of cyclin D1 in Barrett's esophagus was associated

with increased risk of adenocarcinoma *(85)*. In particular, cyclin D1 was overexpressed in both esophageal SCC (71%) and adenocarcinoma (64%). The increased cyclin D1 expression in adenocarcinoma was especially frequent in the intestinal-type lesions *(86)*. Increased cyclin D1 expression and reduced E-cadherin expression were significant prognostic factors in patients with esophageal cancers *(87)*. Overexpression of cyclin D1 also correlated with lymph node metastasis and lymphatic vessel invasion as well as the resistance to chemotherapy *(88)*. Furthermore, a case-control analysis of biopsy specimens obtained at recruitment revealed a statistically significantly increased risk of progression to adenocarcinoma in Barrett's esophagus patients whose biopsy specimens were cyclin D1 positive; thus, cyclin D1-positive staining could be a useful biomarker in identifying Barrett's esophagus patients at high risk of esophageal adenocarcinoma *(85)*. Overall, cyclin D1 overexpression starts early in dysplasia and plays an important in esophageal cancer progression, which is useful for the prediction of patient survival and disease recurrence *(89)*.

3.4. EGFR Overexpression and Esophageal Cancer

EGFR, a tyrosine kinase receptor, plays an important role in cell growth and tumorigenesis. Through binding of its extracellular domain to several different ligands (such as epidermal growth factor [EGF] and transforming growth factor-β), EGFR forms a dimerized receptor and in turn activates its intracellular domain tyrosine kinase (TK) and triggers a cascade of phosphorylation events in the cytoplasm, leading to activation of gene expression and cell proliferation. Both MAPK and AP-1 are downstream cascade genes in the EGFR-signaling pathway. EGFR activates one of the MAPKs (mitogen-activated protein kinase kinase) via Ras and then activates ERK1/2, which in turn translocates into the nucleus and promotes expression of genes such as c-Jun and c-Fos (both of which are part of the AP-1 gene family) and COX-2 expression *(19)*. Several studies have demonstrated that amplification and up-regulation of the EGFR gene and protein have been found in both esophageal SCC and adenocarcinoma as well as their premalignant lesions *(1–5)*. Overexpression of EGFR was significantly correlated with depth of invasion of the tumor *(90)*. EGFR gene amplification may be a useful biological marker for the prediction of lymph node metastasis and of a poorer prognosis for esophageal SCC *(91)*. Moreover, EGFR mutations in esophageal carcinoma are rare but do exist *(92)*.

Furthermore, it is well documented that retinoic acid can suppress expression of EGFR and cyclin D1 *(93)*. Retinoic acids suppress EGF-associated cell proliferation through inhibition of EGFR-dependent ERK1/2 activation *(94)*. Immortalized human bronchial epithelial cells can be malignantly transformed by the tobacco carcinogen 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone, which is correlated with overexpression of EGFR

and cyclin D1, whereas retinoid treatment can prevent this transformation with down-regulation of EGFR and cyclin D1 *(93)*. The effects of retinoids are mediated mainly by retinoid receptors (RARs and retinoid X receptors) *(18,19)*. Indeed, our data showed that induction of RAR-β_2 inhibited EGFR expression in esophageal cancer cells and that esophageal cancer risk factors BPDE and bile acid suppressed RAR-β_2 and induced EGFR expression *(19*; our unpublished data). In addition, a study showed that EGFR overexpression in primary esophageal keratinocytes resulted in the biochemical activation of Akt and signal transducer and activator of transcription pathways and induced enhanced cell migration and cell aggregation *(95)*. Finally, targeting of EGFR by using its inhibitors was used for the treatment of esophageal cancer. For example, inhibition of EGFR-TK by erlotinib seems to be a promising novel approach for innovative treatment strategies for esophageal cancer, because it powerfully induced growth inhibition and cell cycle arrest in human esophageal cancer cells and enhanced the antineoplastic effects of other targeted agents *(96)*. Another EGFR-TK inhibitor, gefitinib, has a modest activity in second-line treatment of advanced esophageal cancer *(97)*.

3.5. E-cadherin and Esophageal Cancer

E-cadherin belongs to the cadherin family of Ca^{2+}-dependent cell-cell adhesion molecules, which induce and maintain intercellular connections *(98)*. E-cadherin is the main adhesion molecule of normal epithelia, and it has been implicated in carcinogenesis because its expression is frequently lost in human epithelial cancers *(98)*. In particular, reduced levels of E-cadherin protein not only have been observed in Barrett's dysplasia and esophageal adenocarcinoma *(98,99)* but also have been correlated with poor prognosis *(100)*. The tumors with reduced E-cadherin expression invaded deeper, had more lymph node metastasis, and had more lymphatic invasion than the tumors with preserved E-cadherin expression in esophageal SCC *(101)*. Furthermore, there was a significant decrease in expression of E-cadherin from columnar metaplasia to dysplasia to esophageal adenocarcinoma. However, expression of E-cadherin in columnar metaplasia without dysplasia was similar to that seen in normal squamous epithelium of the esophagus *(102)*. Other studies found that reduced E-cadherin expression detected by immunohistochemistry was correlated with hypermethylation of the E-cadherin gene promoter *(103,104)*. Methylation of multiple genes, including E-cadherin in esophageal adenocarcinoma, was an independent and strong predictor for disease-specific recurrence and survival of patients *(104)*. Moreover, multivariate analysis confirmed that epigenetic tumor profiling by methylation status was a more powerful biomarker of risk prediction in esophageal cancer than the classic clinicopathological features of stage and age *(103)*.

3.6. COX-2 and Esophageal Carcinogenesis

Cyclooxygenase is the key enzyme in arachidonic acid metabolism, and it catalyzes the biosynthesis of prostaglandin H_2, the precursor for prostanoids. COX-2 enzyme is frequently overexpressed in different human cancers, including esophageal cancer, and it is inducible by various agents such as growth factors and tumor promoters (see review in ref. *105)*. However, another isozyme, COX-1, is constitutively expressed in many tissues, and it is thought to be involved in the homeostasis of various physiologic functions in the human body. Increased evidence for possible involvement of COX-2 in esophageal carcinogenesis has been derived from pharmacological analyses of prostaglandins. Various animal and human esophageal tissues have been reported to contain high levels of prostaglandins in cancer *(105)*. Indeed, ex vivo studies demonstrated that COX-2 is overexpressed in esophageal SCC and adenocarcinoma tissues and their premalignant lesions *(105–107)*. A recent study also indicated that both COX isoforms may be involved in the pathogenesis of esophageal adenocarcinoma, because they are linked to the expression of important modulators of angiogenesis (VEGF-A) and lymphangiogenesis (VEGF-C) in esophageal cancer *(108)*. COX-2 up-regulation was associated independently with reduced survival after surgery compared with low COX-2-expressing tumors *(109)*. Overall, the contribution of COX-2 to carcinogenesis and the malignant phenotype of tumor cells has been thought to be related to its abilities to 1) increase production of prostaglandins, 2) convert procarcinogens to carcinogens, 3) inhibit apoptosis, 4) promote angiogenesis, 5) modulate inflammation and immune function, and 6) increase tumor cell invasiveness (see review in ref. *105)*.

3.7. Lost Expression of RAR-β_2 in Esophageal Cancer

Many studies have demonstrated that RAR-β_2 expression is frequently and progressively lost in premalignant and malignant tissues of the esophagus (see reviews in ref. *41)*. Many esophageal, lung, and breast cancer cell lines that do not express RAR-β_2 are resistant to retinoid treatment *(41)*. Restoration of RAR-β_2 expression suppressed esophageal SCC cell growth, induced apoptosis, and inhibited tumor formation in vitro and in nude mice *(18,19)*. Although the molecular mechanisms of RAR-β_2–mediated antitumor effects are not fully understood *(41)*, we found that RAR-β_2–restored sensitivity to retinoic acid was associated with suppression of EGFR, ERK1/2, AP-1, and COX-2 expression in esophageal cancer cells *(19)*.

Our additional data demonstrated that the reduced expression of RAR-β_2 correlated with the increased RAR-β_4 expression in esophageal cancer tissues *(110)*. Induction of RAR-β_4 expression enhances the growth of cancer cells that do not express RAR-β_2 (our unpublished data). Mechanistically, methylation of CpG islands of RAR-β_2 gene promoter is a major cause for lost expression of RAR-β_2 in different cancers *(41)*. BPDE and bile acid

suppressed RAR-β_2 expression in premalignant and malignant esophageal cells, and methylation of RAR-β_2 gene promoter may be responsible *(17,19)*. Together, these studies indicated that RAR-β_2 plays an important role in suppressing cancer development, suggesting that RAR-β_2 is a tumor-suppressor gene (see reviews in **ref**. *41)*.

3.8. RRIG1 Antitumor Activities in Esophageal Cancer

By using restriction fragment differential display-polymerase chain reaction, we cloned a novel retinoid receptor-induced gene, named RRIG1, which is differentially expressed by RAR-β_2–positive and RAR-β_2–negative esophageal SCC cells. RRIG1 is expressed in a broad range of normal tissues, but it is lost in various cancer specimens, including esophageal SCC and adenocarcinoma *(111,112)*. BPDE was able to suppress RRIG1 expression in esophageal cancer cell lines. RRIG1 mediates the effect of RAR-β_2 on cell growth and gene expression (e.g., ERK1/2 and COX-2). The RRIG1 protein is expressed in the cell membrane and binds to and inhibits the activity of the small GTPase RhoA *(111)*. Whereas induction of RRIG1 expression inhibits RhoA activation and F-actin formation and consequently reduces the colony formation, invasion, and proliferation of esophageal cancer cells, antisense RRIG1 increases RhoA activity and F-actin formation and thus induces the colony formation, invasion, and proliferation of these cells. Our findings revealed a novel molecular pathway involving RAR-β_2 regulation of RRIG1 expression and RRIG1–RhoA interaction *(111)*.

Furthermore, RRIG1 expression correlated positively with tumor differentiation but inversely with lymph node metastasis of esophageal SCC. The stable transfection of RRIG1 inhibited esophageal SCC cell growth and the expression of ERK1/2 and cyclin D1. RRIG1-transfected sublines also inhibited tumor development in nude mice. This study indicated that RRIG1 plays a role in suppressing tumorigenesis *(112)*. Further studies are needed to determine the molecular mechanisms responsible for the RRIG-1–suppressed growth and invasiveness of breast and esophageal cancer cells in vitro and in vivo.

3.9. Altered Expression of DNA Repair Genes in Esophageal Cancer

To date, few studies have reported alternation of DNA repair gene expressions (such as ATM) in esophageal cancer. We found that BPDE was able to bind to the ATM gene after we treated the normal esophageal epithelial cells with BPDE *(35)*. Our further study showed that tissues from the esophageal SCC patients who smoked tobacco expressed more ATM than tissues from nonsmokers. BPDE increased ATM protein levels in cultured esophageal cancer cells. Moreover, BPDE increased DNA double-chain breakage in esophageal cancer cells. This study indicates that tobacco smoke can induce DNA damage and expression of the DNA repair gene ATM *(113)*. Others have reported, by using

array-based comparative genomic hybridization analysis, that Barrett's adenocarcinoma frequently gains DNA sequence copy number, including the ATM gene *(114)*. ATM gene mutations contributed to an unusual radiosensitivity of esophageal cancer cells *(115)*.

Furthermore, several studies have demonstrated a frequent LOH in the region of the BRCA1 gene locus on chromosome 17q in Barrett's esophagus and in esophageal SCC and adenocarcinoma *(116,117)*, suggesting that BRCA1 may be a candidate tumor-suppressor gene in esophageal cancer. In this perspective, our recent data demonstrated that BPDE was able to bind to the BRCA1 gene after the normal esophageal epithelial cells were treated with BPDE *(35)*, which indicates that further study is needed to investigate whether these two genes play a role in the suppression of esophageal carcinogenesis.

4. Chemoprevention of Esophageal Cancer

Many of preclinical in vitro and in vivo studies have been aimed at the chemoprevention of esophageal cancer *(118–124)*. Some agents were shown to inhibit tumor initiation by suspected human carcinogens in animal models. Others showed in vitro activity in cell lines from esophageal cancers. These agents include epigallocatechin-3-gallate (EGCG), bestatin, curcumin, black raspberries, 5-lipoxygenase (LOX), COX-2 inhibitors, difluoromethylornithine, diallyl sulfide, isothiocyanates, and several polyphenolic compounds *(50,118–124)*. This section discusses the most frequently used chemopreventive agents to date, which could be explored in the future for drug combination study to target molecular pathways in esophageal cancer (*see* **Fig. 17.1**).

4.1. Retinoic Acids in the Prevention of Esophageal Cancer

The retinoids were the first group of agents used in the clinic for the prevention of esophageal cancer *(125)*. Retinoids are a group of structural and functional analogs of vitamin A that are required for the maintenance of normal cell growth and differentiation in epithelial tissues *(41)*. The rationale to use retinoids was based on the fact that vitamin A deficiency was found in esophageal cancer patients, and its deficiency induced hyperkeratotic change in the esophageal mucosa in experimental animals *(125)*. Previous clinical chemoprevention studies using retinoic acid showed efficacy in preventing head and neck, lung, and skin cancers *(125)*. Moreover, several studies showed the particular importance of retinoic acid in the suppression of EGFR or cyclin D1 expression *(93,94)*. All-*trans*-retinoic acid and *N*-(4-hydroxyphenyl)retinamide induce morphological changes and inhibit growth through down-regulation

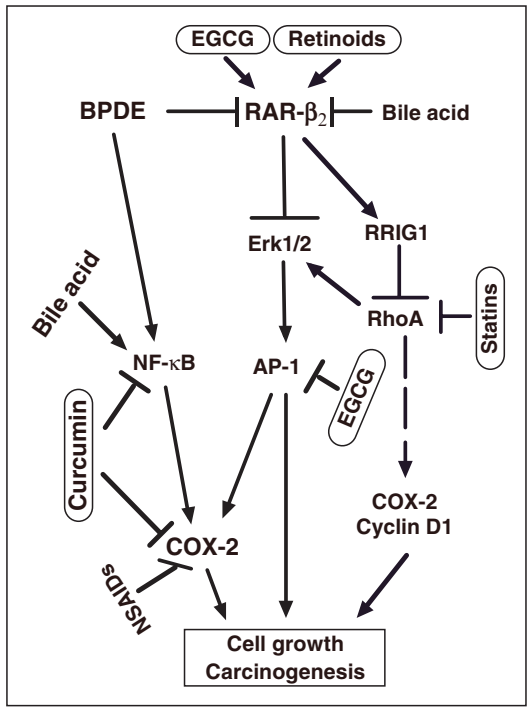


Fig. 17.1. Targeting the RAR-β_2–led molecular pathway as one strategy in the prevention of esophageal cancer. BPDE and bile acid are risk factors for esophageal carcinogenesis. They are believed to exert their effects through the suppression of RAR-β_2 expression and, in turn, up-regulation of EGFR expression and ERK1/2 phosphorylation, resulting in the induction of AP-1 and COX-2 expression. Furthermore, BPDE and bile acid can activate NF-κB to induce greater COX-2 expression. In addition, BPDE can reduce RRIG1 expression (perhaps through reduction of RAR-β_2). RRIG1 mediates the effects of RAR-β_2 on regulating cancer cell growth and gene expression. RRIG1 protein binds to and inhibits RhoA activity and consequently suppresses ERK1/2 phosphorylation and expression of COX-2 and cyclin D1, resulting in reductions in tumor cell colony formation, invasion, and proliferation. Therefore, targeting this gene pathway by using the combination of EGCG, curcumin, and statins may effectively prevent esophageal cancer in both in vitro and in vivo studies.

of EGFR *(126)*. Other studies showed inhibition of cyclin D1 expression by retinoids via their nuclear receptors *(93)*. Indeed, a clinical trial using *N*-4-(ethoxycarbophenyl)retinamide demonstrated that cancer incidence among the treatment group with severe esophageal dysplasia was reduced by 43.2% compared with the placebo group *(127)*. However, the results from other two trials conducted in Linxian, China, by the National Cancer Institute were inconclusive *(128,129)*. Our in vitro study demonstrated that esophageal cancer cells that do not express RAR-β_2 are resistant to retinoic acid *(41)*. In animal experiments, dietary *N*-(4-hydroxyphenyl)retinamide enhanced tumorigenesis in response to *N*-nitrosomethylbenzylamine in the rat esophagus by increasing tumor initiation events *(130)*. Again, 13-*cis*-retinoic acid could not reduce NBMA-induced esophageal tumor multiplicity in rats *(131)*. Furthermore, a randomized clinical trial showed that 13-*cis*-retinoic acid treatment did not reduce the

overall rates of second primary tumors, recurrences, or mortality in patients with stage I non-small cell lung cancer, and secondary multivariate and subset analyses suggested that 13-*cis*-retinoic acid was harmful in current smokers and beneficial in never-smokers *(132)*. Another clinical trial in head and neck cancer also showed that retinoic acid failed to prevent second primary tumors *(133)*. Thus, these studies indicated that a novel and effective approach is needed in the prevention of esophageal cancer.

4.2. Nonsteroidal Anti-Inflammatory Drugs (NSAIDs) in Prevention of Esophageal Cancer

NSAIDs, including aspirin and its related drugs, are the most widely used medicines around the world. Epidemiologic, experimental, and clinical studies have demonstrated that aspirin and some other NSAIDs can decrease the incidence of esophageal cancer among regular users or in experimental animals *(105)*. Several studies, including ours, showed that NSAIDs could induce esophageal cancer cells to undergo apoptosis, which is associated with COX-2 expression *(105)*. Our data showed that while inducing apoptosis in esophageal cancer cells, aspirin or NS398 induced the release of cytochrome *c* from mitochondria, activating caspase-9 and caspase-3 and resulting in the cleavage of poly(ADP-ribose) polymerase *(105)*. The effects of these NSAIDs were associated with their abilities to inhibit COX-2 enzymatic activity and to up-regulate the expression of 15-LOX-1 and -2 *(105)*. In animal models, treatment with selective COX-2 inhibitors has been shown to reduce the damage caused by acid and pepsin in rabbit esophageal mucosa and the incidence of esophageal adenocarcinoma in an animal model and to induce apoptosis in both Barrett's epithelial and adenocarcinoma cells *(134)*. Chemopreventive treatments with NSAIDs were effective in inhibiting tumor development in NMBA-treated animals only when they reduced prostaglandin E_2 (PGE_2) levels in preneoplastic esophageal tissues approximately to those levels found in normal esophagus *(135)*. Celecoxib was effective in preventing columnar-lined epithelium and esophageal adenocarcinoma by suppressing esophagitis in a rat esophageal adenocarcinoma model with a duodenoesophageal reflux procedure *(136)*. A clinical study has shown that selective inhibition of COX-2 significantly decreased cell proliferation in Barrett's metaplasia *(134)*. The use of NSAIDs might be an effective chemopreventive strategy, reducing the risk of neoplastic progression in Barrett's esophagus *(137)*. (Neo)adjuvant therapy with celecoxib might have clinical potential for patients with esophageal adenocarcinoma *(138)*. However, it has recently been reported that drugs such as rofecoxib (Vioxx) and celecoxib (Celebrex) induced significant side effects, such as cardiovascular events, during a clinical chemoprevention trial against colorectal cancer, which raised safety concerns for high-dose and long-term use of these drugs in cancer prevention studies.

4.3. Green Tea or EGCG in the Prevention of Esophageal Cancer

EGCG is a major water-soluble component of green tea, which shows chemopreventive effects against chemical carcinogenesis at many organ sites, including the forestomach and esophagus in rodents *(124,139)*. Tea consumption also has been shown to prevent human esophageal cancer in an epidemiological study *(124,139)*. In addition, the polyphenols in green tea show antimutagenic activity in vitro and in vivo *(124,139)*. EGCG and its analogs possess antioxidative activities, modulate xenobiotic metabolite enzymes, and inhibit tumor promotion and progression *(124,139)*. Mechanically, EGCG possesses antioxidant/pro-oxidant activity, inhibits growth factor signaling (e.g., EGFR and ERK1/2), and it inhibits the activity of secreted matrix metalloproteinase (MMP)2 and MMP9 (see reviews in refs. *124,139,140)*. A recent study demonstrated that EGCG induced a concentration- and time-dependent reversal of hypermethylation of RAR-β_2 in esophageal cancer cell lines, resulting in re-expression of RAR-β_2 *(34)*. EGCG can significantly inhibit NMBA-induced rat esophageal carcinogenesis, and its inhibitory effects might partly target cyclin D1 and COX-2 expression and PGE_2 production *(141)*.

4.4. Curcumin in Prevention of Human Cancers

Curcumin is a polyphenol derived from the roots of *Curcumin longa*, and it is used in Indian cooking. In vitro and in vivo studies have shown that it can inhibit tumor initiation induced by tobacco smoke, benzo[a]pyrene, and 7,12-dimethylbenz[a]anthr acene and tumor promotion induced by phorbol ester *(124,142)*. It also showed the ability to suppress growth and induce apoptosis in numerous types of cancer cells in vitro *(124,142)*. Although the defined mechanisms of its action require further study, its efficacy seems to be related to the induction of glutathione and glutathione transferase activity, inhibition of lipid peroxidation and arachidonic acid metabolism, and suppression of oxidative DNA adduct formation *(124,142,143)*. Curcumin is a potent inhibitor of protein kinase C, EGFR kinase, and IκB kinase *(144)*. Curcumin can inhibit the activation of NF-κB and the expression of c-Jun, c-Fos, c-Myc, ERK1/2, COX-2, phosphatidylinositol 3-kinase (PI3K), Akt, CDKs, and iNOS *(143,144)*. Curcumin also was able to suppress cigarette smoke-induced NF-κB activation and COX-2 expression in head and neck SCC and non-small-cell lung cancer cells *(145,146)*. In esophageal cancer, dietary curcumin can inhibit chemically induced esophageal carcinogenesis in mice and rats *(119,147)*.

4.5. Statins in Prevention of Human Cancers

The statin family of drugs is composed of low-molecular-weight inhibitors of the rate-limiting enzyme of the mevalonate pathway, 3-hydroxy-3-methylglutaryl-CoA reductase (see reviews in refs. *148*, *149)*. Statins have been approved and are effectively used in the control of hypercholesterolemia. Recent evidence,

however, also showed that they have cancer-chemopreventive effects *(148,149)*. Statins can trigger some tumor cells to undergo apoptosis in vitro and suppress tumor growth in vivo. Statins also have an antimetastatic property, which is evident in their suppression of the invasiveness of tumor cells in Matrigel, as well as in animal experiments *(148,149)*. In addition, statins, especially at high concentrations, can inhibit capillary tube formation by endothelial cells in vitro and in vivo *(150)*. The effects of statins are thought to be mediated through the inhibition of Ras and RhoA activity *(148,149)*, although inhibition of the PI3K pathway and cell cycle progression also may also be the target of statins *(148,149)*. Our unpublished data also confirmed RhoA inhibition by statins to induce apoptosis and cell growth arrest in prostate cancer cells.

4.6. Use of Drug Combinations in Prevention of Human Cancers

At the present time, we face at least two kinds of difficulties in clinical chemoprevention studies. First, because the causes of carcinogenesis and the carcinogenic process in human cells remain to be defined, it is difficult for us to select a special molecular target for developing chemopreventive drugs **(2)**. For example, treatment with a single agent such as retinoic acid for the prevention or delay of carcinogenesis was not that effective. Second, the high-dose and long-term use of certain drugs such as rofecoxib and celecoxib may cause unwanted side effects in chemoprevention of human cancer. Therefore, the use of a drug combination may be more practical and effective in clinical chemoprevention studies. The first major combination treatment was reported by Torrance et al. *(151)*, in which mouse colorectal adenomas were prevented by using the combination of COX-2 and EGFR inhibitors, and the synergistic effect was achieved in vitro due to the convergence of these two signaling pathways. Therefore, a combination of statin, retinoid, and curcumin or EGCG will have additive or even synergistic effects against esophageal cancer **(Fig.17.1).** Indeed, a previous study suggested that the combination of statin with retinoid might be a useful anticancer strategy for the prevention of human cancers *(152)*.

5. Future Directions

Esophageal cancer remains an aggressive and lethal disease, and the prognosis is worse than that of many other cancers despite some advancement in its diagnosis and multimodality treatment *(1–5,7)*. Moreover, the incidence of esophageal adenocarcinoma has risen rapidly over the past 30 years in the United States and in several Western European countries *(1–5,7)*. Therefore,

there is an urgent need to improve this situation. For conquest of this now deadly disease, researchers need to discover novel biomarkers for identifying the high-risk population. Microarray, proteomic, and molecular epidemiologic studies on biomarkers will identify more molecular signatures of the tumor, distinguish esophageal SCC and adenocarcinoma, and predict malignant progression and patient prognosis. They also will help us place genes into their molecular pathways for the mechanistic study of esophageal cancer development and progression. Furthermore, mechanistic studies of the risk factors and esophageal carcinogenesis will show how the risk factors induce gene alteration and the underlying molecular mechanisms for esophageal cancer development. In addition, prevention is also an important goal before a curable therapy is available. Several preventive approaches that are easily implemented should be taken to reduce esophageal cancer incidence, such as changes in lifestyle (such as tobacco smoking cessation, limitation of alcohol intake, and body weight watch) and improved nutrition (such as elimination of pickled vegetable intake and consumption of more fruits and vegetables). Another feasible approach is chemoprevention, which uses nontoxic or less-toxic drugs in preventing or delaying cancer incidence in high-risk populations (e.g., Barrett's esophagus with dysplastic lesions or GERD). However, novel and effective agents need to be explored and discovered. Finally, more precise personalized risk predication and therapy with effective molecular biomarker analysis in the clinic, if available, also should help detect early and curable lesions to reduce the incidence and mortality of esophageal cancer.

Acknowledgments

We thank the Department of Scientific Publications at The University of Texas M. D. Anderson Cancer Center for editing the manuscript. The original work was supported by National Cancer Institute grants R29 CA74835, R21 CA10226, and R01 CA117895 and by the Jerry and Maury Rubenstein Foundation.

References

1. Mandard, A. M., Hainaut, P. and Hollstein, M. (2000) Genetic steps in the development of squamous cell carcinoma of the esophagus. *Mutat. Res.* **462**, 335–342.
2. Stoner, G. D. and Gupta, A. (2001) Etiology and chemoprevention of esophageal squamous cell carcinoma. *Carcinogenesis* **22**, 1737–1746.
3. Chen, X. and Yang, C. S. (2001) Esophageal adenocarcinoma: a review and perspectives on the mechanism of carcinogenesis and chemoprevention. *Carcinogenesis* **22**, 1119–1129.

4. Reid, B. J., Blount, P. L., and Rabinovitch, P. S. (2003) Biomarkers in Barrett's esophagus. *Gastrointest. Endosc. Clin. N. Am.* **13**, 369–397.
5. Spechler, S. J. (2005) Barrett's esophagus: a molecular perspective. *Curr. Gastroenterol. Rep.* **7**, 177–181.
6. Jemal, A., Siegel, R., Ward, E., Murray, T., Xu, J., and Thun, M. J. (2007) Cancer statistics, 2007. *CA Cancer J. Clin.* **57**, 43–66.
7. Yamada, T., Alpers, D. H., Laine, L., Owyang, C., and Powell, D.W. (eds.) (1999) *Textbook of Gastroenterology. Vol. 1, 3rd edition*. Lippincott Williams & Wilkins, Philadelphia, New York, and Baltimore.
8. Blot, W. (1994) Esophageal cancer trends and risk factors. *Semin. Oncol.* **21**, 403–410.
9. van Soest, E. M., Dieleman, J. P., Siersema, P. D., Sturkenboom, M. C., and Kuipers, E. J. (2005) Increasing incidence of Barrett's oesophagus in the general population. *Gut* **54**, 1062–1066.
10. Baquet, C. R., Commiskey, P., Mack, K., Meltzer, S., and Mishra, S. I. (2005) Esophageal cancer epidemiology in blacks and whites: racial and gender disparities in incidence, mortality, survival rates and histology. *J. Natl. Med. Assoc.* **97**, 1471–1478.
11. Olliver, J. R., Hardie, L. J., Gong, Y., Dexter, S., Chalmers, D., Harris, K. M., and Wild, C. P. (2005) Risk factors, DNA damage, and disease progression in Barrett's esophagus. *Cancer Epidemiol. Biomarkers Prev.* **14**, 620–625.
12. Vaughan, T. L., Kristal, A. R., Blount, P. L., Levine, D. S., Galipeau, P. C., Prevo, L. J., Sanchez, C. A., Rabinovitch, P. S., and Reid, B. J. (2002) Nonsteroidal anti-inflammatory drug use, body mass index, and anthropometry in relation to genetic and flow cytometric abnormalities in Barrett's esophagus. *Cancer Epidemiol. Biomarkers Prev.* **11,** 745–752.
13. Smith, K. J., O'Brien, S. M., Smithers, B. M., Gotley, D. C., Webb, P. M., Green, A. C., and Whiteman, D. C. (2005) Interactions among smoking, obesity, and symptoms of acid reflux in Barrett's esophagus. *Cancer Epidemiol. Biomarkers Prev.* **14**, 2481–2486.
14. McManus, D. T., Olaru, A., and Meltzer, S. J. (2004) Biomarkers of esophageal adenocarcinoma and Barrett's esophagus. *Cancer Res.* **64**, 1561–1569.
15. Montesano, R., Hollstein, M., and Hainaut, P. (1996) Genetic alterations in esophageal cancer and their relevance to etiology and pathogenesis: a review. *Int. J. Cancer* **69**, 225–235.
16. Denissenko, M. F., Pao, A., Tang, M., and Pfeifer, G. P. (1996) Preferential formation of benzo[a]pyrene adducts at lung cancer mutational hotspots in p53. *Science* **274**, 430–432.
17. Song, S. and Xu, X. C. (2001) Effect of benzo[a]pyrene diol epoxide on expression of retinoic acid receptor-beta in immortalized esophageal epithelial cells and esophageal cancer cells. *Biochem. Biophys. Res. Commun.* **281**, 872–877.
18. Li, M., Song, S., Lippman, S. M., Zhang, X. K., Liu, X., Lotan, R., and Xu, X. C. (2002) Induction of retinoic acid receptor-β suppresses cyclooxygenase-2 expression in esophageal cancer cells. *Oncogene* **21**, 411–418.
19. Song, S., Lippman, S. M., Zou, Y., Ye, X., Ajani, J. A., and Xu, X.-C. (2005) Induction of cyclooxygenase-2 by benzo[a]pyrene diol epoxide through inhibition of retinoic acid receptor-β_2 expression. *Oncogene* **24**, 8268–8276.
20. Mariette, C., Finzi, L., Piessen, G., Van Seuningen, I., and Triboulet, J. P. (2005) Esophageal carcinoma: prognostic differences between squamous cell carcinoma and adenocarcinoma. *World J. Surg.* **29**, 39–45.
21. U.S. Department of Health and Human Services (1990) Smoking, tobacco, and cancer program 1985-1989 status report. NIH Publ. No. 90-3107.
22. Zhang, Z.F., Kurtz, R.C., Sun, M., Karpeh, M., Jr., Yu, G. P., Gargon, N., Fein, J. S., Georgopoulos, S. K., and Harlap, S. (1996) Adenocarcinomas of the esophagus and gastric cardia: medical conditions, tobacco, alcohol, and socioeconomic factors. *Cancer Epidemiol. Biomarkers Prev.* **5**, 761–768.
23. Gammon, M.D., Schoenberg, J.B., Ahsan, H., Risch, H. A., Vaughan, T. L., Chow, W. H., Rotterdam, H., West, A. B., Dubrow, R., Stanford, J. L., Mayne, S. T., Farrow, D. C., Niwa, S., Blot, W. J., and Fraumeni, J. F., Jr. (1997) Tobacco, alcohol, and socioeconomic status and adenocarcinomas of the esophagus and gastric cardia. *J. Natl. Cancer Inst.* **89**, 1277–1284.
24. Tran, G. D., Sun, X. D., Abnet, C. C., Fan, J. H., Dawsey, S. M., Dong, Z. W., Mark, S. D., Qiao, Y. L., and Taylor, P. R. (2005) Prospective study of risk factors for esophageal and gastric cancers in the Linxian general population trial cohort in China. *Int. J. Cancer* **113**, 456–463.

25. Hoffmann, D. and Hecht, S. S. (1990) Advances in tobacco carcinogenesis. In *Chemical Carcinogenesis and Mutagenesis I. Handbook of Experimental Pharmacology* (Cooper, C. S. and Grover, P. L., eds.), vol. **94**, pp. 63-102. Springer-Verlag, Berlin, Germany.
26. Harvey, R. G. (ed.) (1991) *Polycyclic Aromatic Hydrocarbons: Chemistry and Carcinogenesis*. Cambridge University Press, Cambridge, U.K.
27. Hecht, S. S. (2006) Cigarette smoking: cancer risks, carcinogens, and mechanisms. *Langenbecks Arch. Surg.* **391**, 603–613.
28. Osborne M. R. and Crosby N. T. (eds.) (1987) *Benzopyrenes*. Cambridge University Press, Cambridge, U.K.
29. Bartsch, H. (1996) DNA adducts in human carcinogenesis: etiological relevance and structure-activity relationship. *Mutat. Res.* **340**, 67–79.
30. Venkatachalam, S., Denissenko, M. F., Alvi, N., and Wani, A. A. (1993) Rapid activation of apoptosis in human promyelocytic leukemic cells by (+/−)-anti-benzo[a]pyrene diol epoxide induced DNA damage. *Biochem. Biophys. Res. Commun.* **197**, 722–729.
31. Mass, M. J., Jeffers, A. J., Ross, J. A., Nelson, G., Galati, A. J., Stoner, G. D., and Nesnow, S. (1993) Ki-ras oncogene mutations in tumors and DNA adducts formed by benz[j]aceanthrylene and benzo[a]pyrene in the lungs of strain A/J mice. *Mol. Carcinog.* **8**, 186–192.
32. Wang, X.D., Liu, C., Bronson, R.T., Smith, D.E., Krinsky, N.I., and Russell, M. (1999) Retinoid signaling and activator protein-1 expression in ferrets given beta-carotene supplements and exposed to tobacco smoke. *J. Natl. Cancer Inst.* **91**, 60–66.
33. Vuillemenot, B. R., Pulling, L. C., Palmisano, W. A., Hutt, J. A., and Belinsky, S. A. (2004) Carcinogen exposure differentially modulates RAR-beta promoter hypermethylation, an early and frequent event in mouse lung carcinogenesis. *Carcinogenesis* **25**, 623–629.
34. Fang, M. Z., Wang, Y., Ai, N., Hou, Z., Sun, Y., Lu, H., and Yang, C. S. (2003) Tea polyphenol (−)-epigallocatechin-3-gallate inhibits DNA methyltransferase and reactivates methylation-silenced genes in cancer cell lines. *Cancer Res.* **63**, 7563–7570.
35. Liang, Z., Lippman, S. M., Kawabe, A., Shimada, Y., and Xu, X.-C. (2003) Identification of benzo[a]pyrene diol epoxide-binding DNA fragments using DNA-immunoprecipitation technique. *Cancer Res.* **63**, 1470–1474.
36. Wang, X. D. (2005) Alcohol, vitamin A, and cancer. *Alcohol* **35**, 251–258.
37. Shiraishi-Yokoyama, H., Yokoyama, H., Matsumoto, M., Imaeda, H., and Hibi, T. (2006) Acetaldehyde inhibits the formation of retinoic acid from retinal in the rat esophagus. *Scand. J. Gastroenterol.* **41**, 80–86.
38. Wang, X. D., Liu, C., Chung, J., Stickel, F., Seitz, H. K., and Russell, R. M. (1998) Chronic alcohol intake reduces retinoic acid concentration and enhances AP-1 (c-Jun and c-Fos) expression in rat liver. *Hepatology* **28**, 744–750.
39. Kaneko, M., Morimura, K., Nishikawa, T., Wanibuchi, H., Osugi, H., Kinoshita, H., Koide, A., Mori, Y., and Fukishima, S. (2002) Weak enhancing effects of simultaneous ethanol administration on chemically induced rat esophageal tumorigenesis. *Oncol. Rep.* **9**, 1069–1073.
40. Homann, N., Seitz, H. K., Wang, X. D., Yokoyama, A., Singletary, K. W., and Ishii, H. (2005) Mechanisms in alcohol-associated carcinogenesis. *Alcohol Clin. Exp. Res.* **29**, 1317–1320.
41. Xu, X.-C. (2007) Tumor-suppressive activity of retinoic acid receptor-β in cancer. *Cancer Lett.* **253**, 14–24.
42. Csendes, A., Burdiles, P., Braghetto, I., Smok, G., Castro, C., Korn, O., and Henriquez, A. (2002) Dysplasia and adenocarcinoma after classic antireflux surgery in patients with Barrett's esophagus: the need for long-term subjective and objective follow-up. *Ann. Surg.* **235**, 178–185.
43. Lindblad, M., Rodriguez, L. A., and Lagergren, J. (2005) Body mass, tobacco and alcohol and risk of esophageal, gastric cardia, and gastric non-cardia adenocarcinoma among men and women in a nested case-control study. *Cancer Causes Control* **16**, 285–294.
44. Hampel, H., Abraham, N. S., and El-Serag, H. B. (2005) Meta-analysis: obesity and the risk for gastroesophageal reflux disease and its complications. *Ann. Intern. Med.* **143**, 199–211.
45. Barak, N., Ehrenpreis, E. D., Harrison, J. R., and Sitrin, M. D. (2002) Gastro-oesophageal reflux disease in obesity: pathophysiological and therapeutic considerations. *Obes. Rev.* **3**, 9–15.
46. Triadafilopoulos, G. (2003) Management of Barrett's esophagus with and without dysplasia. *Scand. J. Gastroenterol. Suppl.* **237**, 40–46.
47. Menges, M., Müller, M., and Zeitz, M. (2001) Increased acid and bile reflux in

Barrett's esophagus compared to reflux esophagitis, and effect of proton pump inhibitor therapy. *Am. J. Gastroenterol.* **96**, 331–337.

48. Zhang, F., Altorki, N. K., Wu, Y. C., Soslow, R. A., Subbaramaiah, K., and Dannenberg, A. J. (2001) Duodenal reflux induces cyclooxygenase-2 in the esophageal mucosa of rats: evidence for involvement of bile acids. *Gastroenterology* **121**, 1391–1399.
49. Jaiswal, K., Lopez-Guzman, C., Souza, R. F., Spechler, S. J., and Sarosi, G. A., Jr. (2006) Bile salt exposure increases proliferation through p38 and ERK MAPK pathways in a non-neoplastic Barrett's cell line. *Am. J. Physiol. Gastrointest. Liver Physiol.* **290**, G335–G342.
50. Li, M., Lotan, R., Levin, B., Tahara, E., Lippman, S. M., and Xu, X.-C. (2000) Aspirin induction of apoptosis in esophageal cancer: a potential for chemoprevention. *Cancer Epidemiol. Biomarkers Prev.* **9**, 545–549.
51. Jiang, Z. R., Gong, J., Zhang, Z. N., and Qiao, Z. (2006) Influence of acid and bile acid on ERK activity, PPARgamma expression and cell proliferation in normal human esophageal epithelial cells. *World J. Gastroenterol.* **12**, 2445–2449.
52. Jaiswal, K., Tello, V., Lopez-Guzman, C., Nwariaku, F., Anthony, T., and Sarosi, G. A., Jr. (2004) Bile salt exposure causes phosphatidyl-inositol-3-kinase-mediated proliferation in a Barrett's adenocarcinoma cell line. *Surgery* **136**, 160–168.
53. Jenkins, G. J., Harries, K., Doak, S. H., Wilmes, A., Griffiths, A. P., Baxter, J. N., and Parry, J. M. (2004) The bile acid deoxycholic acid (DCA) at neutral pH activates NF-kappaB and induces IL-8 expression in oesophageal cells in vitro. *Carcinogenesis* **25**, 317–323.
54. Abdel-Latif, M. M., O'Riordan, J., Windle, H. J., Carton, E., Ravi, N., Kelleher, D., and Reynolds, J. V. (2004) NF-kappaB activation in esophageal adenocarcinoma: relationship to Barrett's metaplasia, survival, and response to neoadjuvant chemoradiotherapy. *Ann. Surg.* **239**, 491–500.
55. Stamp, D. H. (2006) Bile acids aided by acid suppression therapy may be associated with the development of esophageal cancers in westernized societies. *Med. Hypotheses* **66**, 154–157.
56. Si, H. X., Tsao, S. W., Poon, C. S., Wong, Y. C., and Cheung, A. L. (2005) Physical status of HPV-16 in esophageal squamous cell carcinoma. *J. Clin. Virol.* **32**, 19–23.
57. Castillo, A., Aguayo, F., Koriyama, C., Torres, M., Carrascal, E., Corvalan, A., Roblero, J. P., Naquira, C., Palma, M., Backhouse, C., Argandona, J., Itoh, T., Shuyama, K., Eizuru, Y., and Akiba, S. (2006) Human papillomavirus in esophageal squamous cell carcinoma in Colombia and Chile. *World J. Gastroenterol.* **12**, 6188–6192.
58. Wu, D. C., Wu, I. C., Lee, J. M., Hsu, H. K., Kao, E. L., Chou, S. H., and Wu, M. T. (2005) *Helicobacter pylori* infection: a protective factor for esophageal squamous cell carcinoma in a Taiwanese population. *Am. J. Gastroenterol.* **100**, 588–593.
59. Kamangar, F., Dawsey, S. M., Blaser, M. J., Perez-Perez, G. I., Pietinen, P., Newschaffer, C. J., Abnet, C. C., Albanes, D., Virtamo, J., and Taylor, P. R. (2006) Opposing risks of gastric cardia and noncardia gastric adenocarcinomas associated with *Helicobacter pylori* seropositivity. *J. Natl. Cancer Inst.* **98**, 1445–1452.
60. Lee, D. H., Anderson, K. E., Folsom, A. R., and Jacobs, D. R., Jr. (2005) Heme iron, zinc and upper digestive tract cancer: the Iowa Women's Health Study. *Int. J. Cancer* **117**, 643–647.
61. Abnet, C. C., Lai, B., Qiao, Y. L., Vogt, S., Luo, X. M., Taylor, P. R., Dong, Z. W., Mark, S. D., and Dawsey, S. M. (2005) Zinc concentration in esophageal biopsy specimens measured by x-ray fluorescence and esophageal cancer risk. *J. Natl. Cancer Inst.* **97**, 301–306.
62. Liu, C. G., Zhang, L., Jiang, Y., Chatterjee, D., Croce, C. M., Huebner, K., and Fong, L. Y. (2005) Modulation of gene expression in precancerous rat esophagus by dietary zinc deficit and replenishment. *Cancer Res.* **65**, 7790–7799.
63. Chino, O., Kijima, H., Shimada, H., Nishi, T., Tanaka, H., Oshiba, G., Kise, Y., Kajiwara, H., Tsuchida, T., Tanaka, M., Tajima, T., and Makuuchi, H. (2000) Clinicopathological studies of esophageal carcinoma in achalasia: analyses of carcinogenesis using histological and immunohistochemical procedures. *Anticancer Res.* **20**, 3717–3722.
64. Rosenberg, P. S., Greene, M. H., and Alter, B. P. (2003) Cancer incidence in persons with Fanconi anemia. *Blood* **101**, 822–826.
65. Koppert, L. B., Wijnhoven, B. P., van Dekken, H., Tilanus, H. W., and Dinjens, W. N. (2005) The molecular biology of esophageal adenocarcinoma. *J. Surg. Oncol.* **92**, 169–190.
66. Lagarde, S. M., Ten Kate, F. J., Richel, D. J., Offerhaus, G. J., and van Lanschot, J. J.

(2007) Molecular prognostic factors in adenocarcinoma of the esophagus and gastroesophageal junction. *Ann. Surg. Oncol.* **14**, 977–991.

67. Hollstein, M., Sidransky, D., Vogelstein, B., and Harris, C. C. (1991) p53 mutations in human cancers. *Science* **253**, 49–53.
68. Reid, B. J. (2001) p53 and neoplastic progression in Barrett's esophagus. *Am. J. Gastroenterol.* **96**, 1321–1323.
69. Shi, S. T., Yang, G. Y., Wang, L. D., Xue, Z., Feng, B., Ding, W., Xing, E. P., and Yang, C. S. (1999) Role of p53 gene mutations in human esophageal carcinogenesis: results from immunohistochemical and mutation analyses of carcinomas and nearby non-cancerous lesions. *Carcinogenesis* **20**, 591–597.
70. Wang, D. Y., Xiang, Y. Y., Tanaka, M., Li, X. R., Li, J. L., Shen, Q., Sugimura, H., and Kino, I. (1994) High prevalence of p53 protein overexpression in patients with esophageal cancer in Linxian, China and its relationship to progression and prognosis. *Cancer* **74**, 3089–3096.
71. Ikeda, G., Isaji, S., Chandra, B., Watanabe, M., and Kawarada, Y. (1999) Prognostic significance of biologic factors in squamous cell carcinoma of the esophagus. *Cancer* **86**, 1396–1405.
72. Kuwahara, M., Hirai, T., Yoshida, K., Yamashita, Y., Hihara, J., Inoue, H., and Toge, T. (1999) p53, p21(Waf1/Cip1) and cyclin D1 protein expression and prognosis in esophageal cancer. *Dis. Esophagus* **12**, 116–119.
73. Shimada, Y., Imamura, M., Shibagaki, I., Tanaka, H., Miyahara, T., Kato, M., and Ishizaki, K. (1997) Genetic alterations in patients with esophageal cancer with short- and long-term survival rates after curative esophagectomy. *Ann. Surg.* **226**, 162–168.
74. Natsugoe, S., Nakashima, S., Matsumoto, M., Xiangming, C., Okumura, H., Kijima, F., Ishigami, S., Takebayashi, Y., Baba, M., Takao, S., and Aikou, T. (1999) Expression of p21WAF1/Cip1 in the p53-dependent pathway is related to prognosis in patients with advanced esophageal carcinoma. *Clin. Cancer Res.* **5**, 2445–2449.
75. Bian, Y. S., Osterheld, M. C., Fontolliet, C., Bosman, F. T., and Benhattar, J. (2002) p16 inactivation by methylation of the CDKN2A promoter occurs early during neoplastic progression in Barrett's esophagus. *Gastroenterology* **122**, 1113–1121.
76. Powell, E. L., Leoni, L. M., Canto, M. I., Forastiere, A. A., Iocobuzio-Donahue, C. A., Wang, J. S., Maitra, A., and Montgomery, E. (2005) Concordant loss of MTAP and p16/CDKN2A expression in gastroesophageal carcinogenesis: evidence of homozygous deletion in esophageal noninvasive precursor lesions and therapeutic implications. *Am. J. Surg. Pathol.* **29**, 1497–1504.
77. Wong, D. J., Paulson, T. G., Prevo, L. J., Galipeau, P. C., Longton, G., Blount, P. L., and Reid, B. J. (2001) p16(INK4a) lesions are common, early abnormalities that undergo clonal expansion in Barrett's metaplastic epithelium. *Cancer Res.* **61**, 8284–8289.
78. Xing, E. P., Nie, Y., Wang, L. D., Yang, G. Y., and Yang, C. S. (1999) Aberrant methylation of p16INK4a and deletion of p15INK4b are frequent events in human esophageal cancer in Linxian, China. *Carcinogenesis* **20**, 77–84.
79. Tokugawa, T., Sugihara, H., Tani, T., and Hattori, T. (2002) Modes of silencing of p16 in development of esophageal squamous cell carcinoma. *Cancer Res.* **62**, 4938–4944.
80. Liu, Q., Yan, Y. X., McClure, M., Nakagawa, H., Fujimura, F., and Rustgi, A. K. (1995) MTS-1 (CDKN2) tumor suppressor gene deletions are a frequent event in esophagus squamous cancer and pancreatic adenocarcinoma cell lines. *Oncogene* **10**, 619–622.
81. Takeuchi, H., Ozawa, S., Shih, C. H., Ando, N., Kitagawa, Y., Ueda, M., and Kitajima, M. (2004) Loss of p16INK4a expression is associated with vascular endothelial growth factor expression in squamous cell carcinoma of the esophagus. *Int. J. Cancer* **109**, 483–490.
82. Fang, M. Z., Chen, D., Sun, Y., Jin, Z., Christman, J. K., and Yang, C. S. (2005) Reversal of hypermethylation and reactivation of p16INK4a, RARbeta, and MGMT genes by genistein and other isoflavones from soy. *Clin. Cancer Res.* **11**, 7033–7041.
83. Shamma, A., Doki, Y., Shiozaki, H., Tsujinaka, T., Yamamoto, M., Inoue, M., Yano, M., and Monden, M. (2000) Cyclin D1 overexpression in esophageal dysplasia: a possible biomarker for carcinogenesis of esophageal squamous cell carcinoma. *Int. J. Oncol.* **16**, 261–266.
84. Shamma, A., Doki, Y., Shiozaki, H., Tsujinaka, T., Inoue, M., Yano, M., Kimura, Y., Yamamoto, M., and Monden, M. (1998) Effect of cyclin D1 and associated proteins on proliferation of esophageal squamous cell carcinoma. *Int. J. Oncol.* **13**, 455–460.

85. Bani-Hani, K., Martin, I. G., Hardie, L. J., Mapstone, N., Briggs, J. A., Forman, D., and Wild, C. P. (2000) Prospective study of cyclin D1 overexpression in Barrett's esophagus: association with increased risk of adenocarcinoma. *J. Natl. Cancer Inst.* **92**, 1316–1321.
86. Arber, N., Gammon, M. D., Hibshoosh, H., Britton, J. A., Zhang, Y., Schonberg, J. B., Roterdam, H., Fabian, I., Holt, P. R., and Weinstein, I. B. (1999) Overexpression of cyclin D1 occurs in both squamous carcinomas and adenocarcinomas of the esophagus and in adenocarcinomas of the stomach. *Hum. Pathol.* **30**, 1087–1092.
87. Imamura, M., Shimada, Y., Ide, H., Kuwano, H., Kato, H., Yamana, H., Shiozaki, H., Kobayashi, S., Sakamoto, T., Nishihara, T., Sasano, H., Mukai, M., Takubo, K., Nakanishi, Y., and Isono, K. (2001) Prognostic significance of CyclinD1 and E-cadherin in patients with esophageal squamous cell carcinoma: multiinstitutional retrospective analysis. Research Committee on Malignancy of Esophageal Cancer, Japanese Society for Esophageal Diseases. *J. Am. Coll. Surg.* **192**, 708–718.
88. Nagasawa, S., Onda, M., Sasajima, K., Makino, H., Yamashita, K., Takubo, K., and Miyashita, M. (2001) Cyclin D1 overexpression as a prognostic factor in patients with esophageal carcinoma. *J. Surg. Oncol.* **78**, 208–214.
89. Shimada, Y., Imamura, M., Watanabe, G., Uchida, S., Harada, H., Makino, T., and Kano, M. (1999) Prognostic factors of oesophageal squamous cell carcinoma from the perspective of molecular biology. *Br. J. Cancer* **80**, 1281–1288.
90. Hanawa, M., Suzuki, S., Dobashi, Y., Yamane, T., Kono, K., Enomoto, N., and Ooi, A. (2006) EGFR protein overexpression and gene amplification in squamous cell carcinomas of the esophagus. *Int. J. Cancer* **118**, 1173–1180.
91. Kitagawa, Y., Ueda, M., Ando, N., Ozawa, S., Shimizu, N., and Kitajima, M. (1996) Further evidence for prognostic significance of epidermal growth factor receptor gene amplification in patients with esophageal squamous cell carcinoma. *Clin. Cancer Res.* **2**, 909–914.
92. Sudo, T., Mimori, K., Nagahara, H., Utsunomiya, T., Fujita, H., Tanaka, Y., Shirouzu, K., Inoue, H., and Mori, M. (2007) Identification of EGFR mutations in esophageal cancer. *Eur. J. Surg. Oncol.* **33**, 44–48.
93. Dragnev, K. H., Petty, W. J., and Dmitrovsky, E. (2003) Retinoid targets in cancer therapy and chemoprevention. *Cancer Biol. Ther.* **2**, S150–156.
94. Lango, M., Wentzel, A. L., Song, J. I., Xi, S., Johnson, D. E., Lamph, W. W., Miller, L., and Grandis, J. R. (2003) Responsiveness to the retinoic acid receptor-selective retinoid LGD1550 correlates with abrogation of transforming growth factor alpha/epidermal growth factor receptor autocrine signaling in head and neck squamous carcinoma cells. *Clin. Cancer Res.* **9**, 4205–4213.
95. Andl, C. D., Mizushima, T., Nakagawa, H., Oyama, K., Harada, H., Chruma, K., Herlyn, M., and Rustgi, A. K. (2003) Epidermal growth factor receptor mediates increased cell proliferation, migration, and aggregation in esophageal keratinocytes in vitro and in vivo. *J. Biol. Chem.* **278,** 1824–1830.
96. Sutter, A. P., Hopfner, M., Huether, A., Maaser, K., and Scherubl, H. (2006) Targeting the epidermal growth factor receptor by erlotinib (Tarceva) for the treatment of esophageal cancer. *Int. J. Cancer* **118**, 1814–1822.
97. Janmaat, M. L., Gallegos-Ruiz, M. I., Rodriguez, J. A., Meijer, G. A., Vervenne, W. L., Richel, D. J., Van Groeningen, C., and Giaccone, G. (2006) Predictive factors for outcome in a phase II study of gefitinib in second-line treatment of advanced esophageal cancer patients. *J. Clin. Oncol.* **24**, 1612–1619.
98. Bongiorno, P. F., al Kasspooles, M., Lee, S. W., Rachwal, W. J., Moore, J. H., Whyte, R. I., Orringer, M. B., and Beer, D. G. (1995) E-Cadherin expression in primary and metastatic thoracic neoplasms and in Barrett's oesophagus. *Br. J. Cancer* **71**, 166–172.
99. Swami, S., Kumble, S., and Triadafilopoulos, G. (1995) E-Cadherin expression in gastroesophageal reflux disease, Barrett's esophagus, and esophageal adenocarcinoma: an immunohistochemical and immunoblot study. *Am. J. Gastroenterol.* **90**, 1808–1813.
100. Krishnadath, K. K., Tilanus, H. W., van Blankenstein, M., Hop, W. C., Kremers, E. D., Dinjens ,W. N., and Bosman, F. T. (1997) Reduced expression of the cadherin-catenin complex in oesophageal adenocarcinoma correlates with poor prognosis. *J. Pathol.* **182**, 331–338.
101. Uchikado, Y., Natsugoe, S., Okumura, H., Setoyama, T., Matsumoto, M., Ishigami, S., and Aikou, T. (2005) Slug expression in the E-cadherin preserved tumors is related

to prognosis in patients with esophageal squamous cell carcinoma. *Clin. Cancer Res.* **11**, 1174–1180.

102. Feith, M., Stein, H. J., Mueller, J., and Siewert, J. R. (2004) Malignant degeneration of Barrett's esophagus: the role of the Ki-67 proliferation fraction, expression of E-cadherin and p53. *Dis. Esophagus* **17**, 322–327.
103. Brock, M. V., Gou, M., Akiyama, Y., Muller, A., Wu, T. T., Montgomery, E., Deasel, M., Germonpre, P., Rubinson, L., Heitmiller, R. F., Yang, S. C., Forastiere, A. A., Baylin, S. B., and Herman, J. G. (2003) Prognostic importance of promoter hypermethylation of multiple genes in esophageal adenocarcinoma. *Clin. Cancer Res.* **9**, 2912–2919.
104. Corn, P. G., Heath, E. I., Heitmiller, R., Fogt, F., Forastiere, A. A., Herman, J. G., and Wu, T. T. (2001) Frequent hypermethylation of the 5′ CpG island of E-cadherin in esophageal adenocarcinoma. *Clin. Cancer Res.* **7**, 2765–2769.
105. Xu, X. C. (2002) COX-2 inhibitors in cancer treatment and prevention, a recent development. *Anticancer Drugs* **13**, 127–137.
106. Zimmermann, K. C., Sarbia, M., Weber, A. A., Borchard, F., Gabbert, H. E., and Schror, K. (1999) Cyclooxygenase-2 expression in human esophageal carcinoma. *Cancer Res.* **59**, 198–204.
107. Wilson, K. T., Fu, S., Ramanujam, K. S., and Meltzer, S. J. (1998) Increased expression of inducible nitric oxide synthase and cyclooxygenase-2 in Barrett's esophagus and associated adenocarcinomas. *Cancer Res.* **58**, 2929–2934.
108. von Rahden, B. H., Stein, H. J., Puhringer, F., Koch, I., Langer, R., Piontek, G., Siewert, J. R., Hofler, H., and Sarbia, M. (2005) Coexpression of cyclooxygenases (COX-1, COX-2) and vascular endothelial growth factors (VEGF-A, VEGF-C) in esophageal adenocarcinoma. *Cancer Res.* **65**, 5038–5044.
109. Heeren, P., Plukker, J., van Dullemen, H., Nap, R., and Hollema, H. (2005) Prognostic role of cyclooxygenase-2 expression in esophageal carcinoma. *Cancer Lett.* **225**, 283–289.
110. Xu, X. C., Lee, J. J., Wu, T. T., Hoque, A., Ajani, J. A., and Lippman, S. M. (2005) Increased retinoic acid receptor-beta4 correlates in vivo with reduced retinoic acid receptor-beta2 in esophageal squamous cell carcinoma. *Cancer Epidemiol. Biomarkers Prev.* **14**, 826–829.
111. Liang, Z. D., Lippman, S. M., Wu, T. T., Lotan, R., Xu, X.-C. (2006) RRIG1 mediates effects of retinoic acid receptor-β_2 on tumor cell growth and gene expression through binding to and inhibiting RhoA. *Cancer Res.* **66**, 7111–7118.
112. Huang, J., Liang, Z. D., Wu, T. T., Hoque, A., Chen, H., Jiang, Y., Zhang, H., and Xu, X.-C. (2007) Tumor-suppressive effect of retinoid receptor-induced gene-1 (RRIG1) in esophageal cancer. *Cancer Res.* **67**, 1589–1593.
113. Jiang, Y., Liang, Z. D., Wu, T. T., Cao, L., Zhang, H., and Xu, X.-C. (2007) ATM expression is associated with tobacco smoke exposure in esophageal cancer tissues and benzo[a]pyrene diol epoxide in cell lines. *Int. J. Cancer* **120**, 91–95.
114. Albrecht, B., Hausmann, M., Zitzelsberger, H., Stein, H., Siewert, J. R., Hopt, U., Langer, R., Hofler, H., Werner, M., and Walch, A. (2004) Array-based comparative genomic hybridization for the detection of DNA sequence copy number changes in Barrett's adenocarcinoma. *J. Pathol.* **203**, 780–788.
115. Ban, S., Michikawa, Y., Ishikawa, K., Sagara, M., Watanabe, K., Shimada, Y., Inazawa, J., and Imai, T. (2005) Radiation sensitivities of 31 human oesophageal squamous cell carcinoma cell lines. *Int. J. Exp. Pathol.* **86**, 231–240.
116. Petty, E. M., Kalikin, L. M., Orringer, M. B., and Beer, D. G. (1998) Distal chromosome 17q loss in Barrett's esophageal and gastric cardia adenocarcinomas: implications for tumorigenesis. *Mol. Carcinog.* **22**, 222–228.
117. Mori, T., Aoki, T., Matsubara, T., Iida, F., Du, X., Nishihira, T., Mori, S., and Nakamura, Y. (1994) Frequent loss of heterozygosity in the region including BRCA1 on chromosome 17q in squamous cell carcinomas of the esophagus. *Cancer Res.* **54**, 1638–1640.
118. Kresty, L. A., Morse, M. A., Morgan, C., Carlton, P. S., Lu, J., Gupta, A., Blackwood, M., and Stoner, G. D. (2001) Chemoprevention of esophageal tumorigenesis by dietary administration of lyophilized black raspberries. *Cancer Res.* **61**, 6112–6119.
119. Huang, M.-T., Lou, Y.-R., Ma, W., Newmark, H. L., Reuhl, K. R., and Conney, A. H. (1994) Inhibitory effects of dietary curcumin on forestomach, duodenal, and colon carcinogenesis in mice. *Cancer Res.* **54**, 5841–5847.

120. Hoque, A., Lippman, S. M., Wu, T. T., Xu, Y., Liang, Z. D., Swisher, S., Zhang, H., Cao, L., Ajani, J. A., and Xu, X.-C. (2005) Increased 5-lipoxygenase expression and induction of apoptosis by its inhibitors in esophageal cancer: a potential target for prevention. *Carcinogenesis* **26**, 785–791.
121. Chen, X., Wang, S., Wu, N., Sood, S., Wang, P., Jin, Z., Beer, D. G., Giordano, T. J., Lin, Y., Shih, W. C., Lubet, R. A., and Yang, C. S. (2004) Overexpression of 5-lipoxygenase in rat and human esophageal adenocarcinoma and inhibitory effects of zileuton and celecoxib on carcinogenesis. *Clin. Cancer Res.* **10**, 6703–6709.
122. Chen, X., Li, N., Wang, S., Wu, N., Hong, J., Jiao, X., Krasna, M. J., Beer, D. G., and Yang, C. S. (2003) Leukotriene A4 hydrolase in rat and human esophageal adenocarcinomas and inhibitory effects of bestatin. *J. Natl. Cancer Inst.* **95**, 1053–1061.
123. Fong, L. Y., Pegg, A. E., and Magee, P. N. (1998) Alpha-difluoromethylornithine inhibits N-nitrosomethylbenzylamine-induced esophageal carcinogenesis in zinc-deficient rats: effects on esophageal cell proliferation and apoptosis. *Cancer Res.* **58**, 5380–5388.
124. Conney, A. H. (2003) Enzyme induction and dietary chemicals as approaches to cancer chemoprevention: the Seventh DeWitt S. Goodman Lecture. *Cancer Res.* **63**, 7005–7031.
125. Sporn, M. B., Roberts, A. B., and Goodman, D. S. (eds.) (1994) *The Retinoids: Biology, Chemistry, and Medicine, 2nd edition*. Raven Press, New York.
126. Muller, A., Nakagawa, H., and Rustgi, A. K. (1997) Retinoic acid and N-(4-hydroxyphenyl)retinamide suppress growth of esophageal squamous carcinoma cell lines. *Cancer Lett.* **113**, 95–101.
127. Han, J. (1993) Highlights of the cancer chemoprevention studies in China. *Prev. Med.* **22**, 712–722.
128. Blot, W. J., Li, J. Y., and Taylor, P. R. (1993) Nutritional intervention trials in Linxian, China: supplementation with specific vitamin/mineral combination, cancer incidence and disease-specific mortality in the general population. *J. Natl. Cancer Inst.* **85**, 1483–1492.
129. Li, J. Y., Taylor, P. R., and Li, B. (1993) Nutrition intervention trials in Linxian, China: multiple vitamin/mineral supplementation, cancer incidence and disease-specific mortality among adults with esophageal dysplasia. *J. Natl. Cancer Inst.* **85**, 1492–1498.
130. Gupta, A., Nines, R., Rodrigo, K. A., Aziz, R. A., Carlton, P. S., Gray, D. L., Steele, V. E., Morse, M. A., and Stoner, G. D. (2001) Effects of dietary *N*-(4-hydroxyphenyl)retinamide on *N*-nitrosomethylbenzylamine metabolism and esophageal tumorigenesis in the Fischer 344 rat. *J. Natl. Cancer Inst.* **93**, 990–998.
131. Daniel, E. M. and Stoner, G. D. (1991) The effects of ellagic acid and 13-cis-retinoic acid on N-nitrosobenzylmethylamine-induced esophageal tumorigenesis in rats. *Cancer Lett.* **56**, 117–124.
132. Lippman, S. M., Lee, J. J., Karp, D. D., Vokes, E. E., Benner, S. E., Goodman, G. E., and Hong, W. K. (2001) Randomized phase III intergroup trial of isotretinoin to prevent second primary tumors in stage I non-small-cell lung cancer. *J. Natl. Cancer Inst.* **93**, 605–618.
133. Khuri, F. R., Lee, J. J., Lippman, S. M., Kim, E. S., Cooper, J. S., Benner, S. E., and Hong, W. K. (2006) Randomized phase III trial of low-dose isotretinoin for prevention of second primary tumors in stage I and II head and neck cancer patients. *J. Natl. Cancer Inst.* **98**, 441–450.
134. Piazuelo, E., Jimenez, P., and Lanas, A. (2003) COX-2 inhibition in esophagitis, Barrett's esophagus and esophageal cancer. *Curr. Pharm. Des.* **9**, 2267–2280.
135. Stoner, G. D., Qin, H., Chen, T., Carlton, P. S., Rose, M. E., Aziz, R. M., and Dixit, R. (2005) The effects of L-748706, a selective cyclooxygenase-2 inhibitor, on N-nitrosomethylbenzylamine-induced rat esophageal tumorigenesis. *Carcinogenesis* **26**, 1590–1595.
136. Oyama, K., Fujimura, T., Ninomiya, I., Miyashita, T., Kinami, S., Fushida, S., Ohta, T., and Koichi, M. (2005) A COX-2 inhibitor prevents the esophageal inflammation-metaplasia-adenocarcinoma sequence in rats. *Carcinogenesis* **26**, 565–570.
137. Vaughan, T. L., Dong, L. M., Blount, P. L., Ayub, K., Odze, R. D., Sanchez, C. A., Rabinovitch, P. S., and Reid, B. J. (2005) Non-steroidal anti-inflammatory drugs and risk of neoplastic progression in Barrett's oesophagus: a prospective study. *Lancet Oncol.* **6**, 945–952.
138. Tuynman, J. B., Buskens, C. J., Kemper, K., ten Kate, F. J., Offerhaus, G. J., Richel, D. J., and van Lanschot, J. J. (2005) Neoadjuvant selective COX-2 inhibition down-regulates important oncogenic pathways in patients

with esophageal adenocarcinoma. *Ann. Surg.* **242**, 840–849, discussion 849–850.

139. Lambert, J. D. and Yang, C. S. (2003) Mechanisms of cancer prevention by tea constituents. *J. Nutr.* **133**, 3262S–3267S.
140. Hou, Z., Sang, S., You, H., Lee, M. J., Hong, J., Chin, K. V., and Yang, C. S. (2005) Mechanism of action of (–)-epigallocatechin-3-gallate: auto-oxidation-dependent inactivation of epidermal growth factor receptor and direct effects on growth inhibition in human esophageal cancer KYSE 150 cells. *Cancer Res.* **65**, 8049–8056.
141. Li, Z. G., Shimada, Y., Sato, F., Maeda, M., Itami, A., Kaganoi, J., Komoto, I., Kawabe, A., and Imamura, M. (2002) Inhibitory effects of epigallocatechin-3-gallate on N-nitrosomethylbenzylamine-induced esophageal tumorigenesis in F344 rats. *Int. J. Oncol.* **21**, 1275–1283.
142. Aggarwal, B. B., Kumar, A., and Bharti, A. C. (2003) Anticancer potential of curcumin: preclinical and clinical studies. *Anticancer Res.* **23**, 363–398.
143. Joe, B., Vijaykumar, M., and Lokesh, B. R. (2004) Biological properties of curcumin–cellular and molecular mechanisms of action. Crit. Rev. *Food Sci. Nutr.* **44**, 97–111.
144. Lin, J. K. (2004) Suppression of protein kinase C and nuclear oncogene expression as possible action mechanisms of cancer chemoprevention by curcumin. *Arch. Pharm. Res.* **27**, 683–692.
145. Aggarwal, S., Takada, Y., Singh, S., Myers, J. N., and Aggarwal, B. B. (2004) Inhibition of growth and survival of human head and neck squamous cell carcinoma cells by curcumin via modulation of nuclear factor-kappaB signaling. *Int. J. Cancer* **111**, 679–692.
146. Shishodia, S., Potdar, P., Gairola, C. G., and Aggarwal, B. B. (2003) Curcumin (diferuloylmethane) down-regulates cigarette smoke-induced NF-kappaB activation through inhibition of IkappaBalpha kinase in human lung epithelial cells: correlation with suppression of COX-2, MMP-9 and cyclin D1. *Carcinogenesis* **24**, 1269–1279.
147. Ushida, J., Sugie, S., Kawabata, K., Pham, Q. V., Tanaka, T., Fujii, K., Takeuchi, H., Ito, Y., and Mori, H. (2000) Chemopreventive effect of curcumin on N-nitrosomethylben-zylamine-induced esophageal carcinogenesis in rats. *Jpn. J. Cancer Res.* **91**, 893–898.
148. Graaf, M. R., Richel, D. J., van Noorden, C. J., and Guchelaar, H. J. (2004) Effects of statins and farnesyltransferase inhibitors on the development and progression of cancer. *Cancer Treat. Rev.* **30**, 609–641.
149. Demierre, M. F., Higgins, P. D., Gruber, S. B., Hawk, E., and Lippman, S. M. (2005) Statins and cancer prevention. *Nat. Rev. Cancer* **5**, 930–942.
150. Vincent, L., Chen, W., Hong, L., Mirshahi, F., Mishal, Z., Mirshahi-Khorassani, T., Vannier, J. P., Soria, J., and Soria, C. (2001) Inhibition of endothelial cell migration by cerivastatin, an HMG-CoA reductase inhibitor: contribution to its anti-angiogenic effect. *FEBS Lett.* **495**, 159–166.
151. Torrance, C. J., Jackson, P. E., Montgomery, E., Kinzler, K. W., Vogelstein, B., Wissner, A., Nunes, M., Frost, P., and Discafani, C. M. (2000) Combinatorial chemoprevention of intestinal neoplasia. *Nat. Med.* **6**, 1024–1028.
152. Kumar, B., Cole, W. C., and Prasad, K. N. (2001) Alpha tocopheryl succinate, retinoic acid and polar carotenoids enhanced the growth-inhibitory effect of a cholesterol-lowering drug on immortalized and transformed nerve cells in culture. *J. Am. Coll. Nutr.* **20**, 628–636.

Chapter 18

Single Nucleotide Polymorphisms in DNA Repair Genes and Prostate Cancer Risk

Jong Y. Park ,Yifan Huang, and Thomas A. Sellers

Summary

The specific causes of prostate cancer are not known. However, multiple etiologic factors, including genetic profile, metabolism of steroid hormones, nutrition, chronic inflammation, family history of prostate cancer, and environmental exposures are thought to play significant roles. Variations in exposure to these risk factors may explain interindividual differences in prostate cancer risk. However, regardless of the precise mechanism(s), a robust DNA repair capacity may mitigate any risks conferred by mutations from these risk factors. Numerous single nucleotide polymorphisms (SNPs) in DNA repair genes have been found, and studies of these SNPs and prostate cancer risk are critical to understanding the response of prostate cells to DNA damage. A few SNPs in DNA repair genes are associated with significantly increased risk of prostate cancer; however, in most cases, the effects are moderate and often depend upon interactions among the risk alleles of several genes in a pathway or with other environmental risk factors. This report reviews the published epidemiologic literature on the association of SNPs in genes involved in DNA repair pathways and prostate cancer risk.

Key words: DNA repair, polymorphism, environmental exposure, prostate cancer, cancer susceptibility.

1. Introduction

Prostate is the most common site of cancer and the third leading cause of cancer mortality in men in the United States *(1)*. There is a large variation in prostate cancer incidence rates among ethnic groups. Incidence rates are the highest among African Americans (272 per 100,000 per year) and the lowest among Asians (2 per 100,000 per year) *(1,2)*. Incidence of prostate cancer is increasing steadily in almost all countries, and the lifetime risk of prostate cancer for men in the United States is 18% *(1,3)*.

Mukesh Verma (ed.), *Methods in Molecular Biology, Cancer Epidemiology, vol. 471*
© 2009 Humana Press, a part of Springer Science + Business Media, Totowa, NJ
Book doi: 10.1007/978-1-59745-416-2

Although the specific causes of prostate cancer are not known, androgens, estrogens, inflammation, and DNA repair capacity have been implicated. Androgens, which play an important part in development and maintenance of the prostate, can induce prostate cancer in rodents *(4)*, and they stimulate the in vitro proliferation of prostate cancer cells *(5)*. Carcinogenesis in prostate tissue involves multiple genetic events.

DNA is constantly damaged by endogenous oxygen free radicals and exogenous chemicals. DNA mutations are estimated to spontaneously occur 20,000–40,000 times every day *(6,7)*. The DNA repair process is important to the survival of the cell; therefore, different repair pathways are available to reverse the different types of DNA damage. In fact, >150 DNA repair enzymes participate in this process *(8)*. Defects in these DNA repair pathways may increase persistent mutations in daughter cell generations, genomic instability, and ultimately prostate cancer risk. These DNA repair genes can be classified into several distinct pathways: direct reversal, base excision repair (BER), nucleotide excision repair (NER), mismatch repair (MMR), and double-strand break repair (DSBR). Depending upon the DNA-damaging agents, different levels of contribution from different classes of DNA repair enzymes could be expected.

In this chapter, we focus on single nucleotide polymorphisms (SNPs) and phenotypes in DNA repair genes that have been investigated in published epidemiologic studies of prostate cancer.

2. Methods

Numerous SNPs in different DNA repair genes have been identified, and many of them have been investigated in relation to human cancer susceptibility *(9)*. We identified studies relevant to prostate cancer using the search engine PubMed (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi) in June 2007. The inclusion criteria for this chapter were epidemiologic studies of the association between polymorphisms in DNA repair genes and prostate cancer risk. Among 24 studies obtained from the search phrases “DNA repair” AND “prostate cancer” AND “polymorphism,” 10 epidemiologic studies were included after review of the articles *(10–19)*. Among 11 additional epidemiologic studies that were obtained after searching by single DNA repair gene name AND “prostate cancer,” five studies were excluded because they reported associations between phenotypes, such as expressions or activities of DNA enzymes and prostate cancer *(20–24)*.

The remaining six studies were included in this chapter *(25–30)*. One article *(19)* was excluded because some of the data published earlier *(30)*; thus, in total 15 published studies form the basis of this review.

The following notation is used to describe SNPs: uppercase letters represent amino acids, with numbers indicating the codon; and lowercase letters represent nucleotides, with numbers indicating the sequence position.

3. Results

By the end of June 2007, associations between SNPs in DNA repair genes and risk of prostate cancer have been reported in 15 published studies. **Table 18.1** provides details on case-control studies of DNA repair gene polymorphisms and levels of association. Most studies were conducted in North America, and four studies were conducted in China *(12,30)*, Japan *(10)*, and United Kingdom *(29)*. Six studies were relatively large (438–996 cases) *(13,17,18,25,27,29)*, but nine studies included 250 or fewer cases. Nine studies were hospital based case-control studies, and four studies were population-based studies *(12,15,26,29)*. Two studies used sibling and family based designs *(13,18)*.

Table 18.2 displays the SNPs in DNA repair genes included in this chapter with allele frequencies, SNP identification number, and their potential functional effects.

3.1. BER Pathway

The BER pathway targets DNA damaged during replication or by environmental agents. Repair of DNA mutations is necessary so that sequence errors are not transmitted to daughter cells. The single damaged base in DNA caused by endogenous metabolism or environmental oxidizing agents result in DNA adducts. This damage has been proposed to play a critical role in carcinogenesis in prostate tissue. Base excision repair involves removing the mutated base out of the DNA and repairing the base alone (**Fig. 18.1A**).

Repair of a mutated base is primarily conducted by enzymes involved in BER with apurinic/apyrimidic (A/P) endonuclease (APE1), human 8-oxoguanine DNA glycosylase (hOGG1), DNA ligase, DNA polymerase delta (POLD1), X-ray repair cross-complementing group 1 (XRCC1), and poly(ADP-ribose) polymerase (ADPRT) *(31–33)*.

3.1.1. Human 8-Oxoguanine DNA Glycosylase (hOGGI)

The enzyme hOGG1 catalyzes the excision and removal of single base adducts *(34,35)*. Base excision repair by hOGG1 enzyme leaves a single nucleotide space. This space is filled by DNA

Table 18.1
Epidemiologic studies of DNA repair gene SNPs and prostate cancer risk

Gene	Race	Study design	Case	Control	Genotype distribution[a]	Comparison	Adjustment[b]	OR (95% CI)	Reference
ADPRT V762A	Caucasian	Hospital	438	427	70,26,4 73,25,2	VA vs. VV AA vs. VV	Age, BPH, FH	1.2 (0.9–1.6) 2.7 (1.1–6..5)	*(27)*
	African Americans	Hospital	50	97	90,10,0 91,9,0	VA vs. VV AA vs. VV		1.0 (0.2–4.0) NA	
APE1 D148E	Caucasian	Hospital	228	219	28,54,18 34,50,16	DE vs. DD EE vs. DD	Age, S, FH	1.2 (0.8–1.9) 1.2 (0.7–2.2)	*(11)*
	African Americans	Population	124	116	34,52,14 38,53,10	DE vs. DD EE vs. DD	Age, S	1.2 (0.7–2.2) 1.6 (0.6–3.8)	
APE1 Q51H	Caucasian	Hospital	228	219	89,10, 1 85,14, 1	QH vs. QQ HH vs. QQ	Age, S, FH	0.6 (0.3–1.2) 0.6 (0.03–14)	*(11)*
	African Americans	Population	124	116	79,19, 2 71,26, 3	QH vs. QQ HH vs. QQ	Age, S	0.8 (0.4–1.5) 0.7 (0.1–3.1)	
ATM D1853N	Caucasians	Population	637	455	73,24,3 69,28,3	ga vs. gg aa vs. gg	None	0.8 (0.7–1.1) 1.0 (0.5–2.1)	*(29)*
ATM D1853V	Caucasians	Population	637	455	99,1,0 99,1,0	ta vs. aa	None	0.9 (0.2–3.3)	*(29)*
ATM ivs38-8t>c	Caucasians	Population	637	455	93,7,0 93,7,0	ct/cc vs. tt	None	1.0 (0.6–1.6)	*(29)*
ATM ivs38-15g>c	Caucasians	Population	637	455	98,2,0 99,1,0	gc vs. gg	None	1.8 (0.6–5.7)	*(29)*
ATMP1054R	Caucasians	Population Hospital	637	455	92,7,0 96,4,0	PR/RR vs. PP	None	2.1 (1.2–3.9)	*(29)*
hOGG1 10660t	Mixed (93% Caucasians)	Hospital	245	222	65,32, 4 61,32, 7	tt vs. aa	Age	NS	*(16)*
hOGG1 a11657g	Mixed (93% Caucasians)	Hospital	245	222	70,25, 5 74,25, 1	gg vs. aa	Age	9.8 (1.2–76.9)	*(16)*
	Mixed (84% Caucasians)	Family	159	222	67,29, 5 74,25, 1	gg vs. aa	Age	13.9 (1.6–125)	

hOGG1 a11657g	Mixed (84% Caucasians)	Hospital	996	1092	70,27, 3 71,25, 4	ag vs. aa gg vs. aa	Age	NS	*(25)*
hOGG1 a11826t	Mixed (93% Caucasians)	Hospital	245	222	66,31, 3 60,33, 7	tt vs. aa	Age	NS	*(16)*
hOGG1 a7143g	Mixed (93% Caucasians)	Hospital	245	222	68,26, 5 71,28, 1	gg vs. aa	Age	5.1 (1.1–23.3)	*(16)*
	Mixed (84% Caucasians)	Family	159	222	64,32, 5 71,28, 1	gg vs. aa	Age	8.2 (1.5–45.5)	
hOGG1 a9110g	Mixed (93% Caucasians)	Hospital	245	222	66,31, 3 60,34, 7	gg vs. aa	Age	NS	*(16)*
hOGG1 c10629g	Mixed (93% Caucasians)	Hospital	245	222	28,44,27 30,41,30	gg vs. cc	Age	NS	*(16)*
hOGG1 g3402a	Mixed (93% Caucasians)	Hospital	245	222	40,44,16 43,45,12	aa vs. gg	Age	NS	*(16)*
hOGG1 g3574a	Mixed (93% Caucasians)	Hospital	245	222	67,28, 5 60,34, 6	aa vs. gg	Age	NS	*(16)*
hOGG1 g6170c	Mixed (93% Caucasians)	Hospital	245	222	63,33, 4 58,34, 7	cc vs. gg	Age	NS	*(16)*
hOGG1 S326C	Caucasians	Hospital	84	252	58,35, 7 74,25, 1	SC vs. SS CC vs. SS	Age, S	1.8 (1.0–3.3) 7.8 (1.7–36)	*(14)*
hOGG1 S326C	Mixed (93% Caucasians)	Hospital	245	222	61,36, 3 55,36, 9	CC vs. SS	Age	0.3 (0.1–0.8)	*(16)*
	Mixed (84% Caucasians)	Family	159	222	61,35, 4 55,36, 9	CC vs. SS	Age	0.5 (0.2–1.7)	
hOGG1 S326C	Mixed (84% Caucasians)	Hospital	996	1092	60,35, 5 57,35, 8	SC vs. SS CC vs. SS	Age, race, FH, DRE, PSA	0.9 (0.8–1.1) 0.7 (0.5–1.0)	*(25)*
hOGG1 S326C	Mixed (90% Caucasians)	Family	439	479	64,31, 5 54,30, 6	SC vs. SS CC vs. SS	Age	1.1 (0.7–1.6) 0.7 (0.3–1.7)	*(18)*
MGMT I143V	Chinese	Population	162	251	96, 3, 1 98, 2, 0	VV/IV vs. II	Age	1.9 (0.6–6.2)	*(12)*

(continued)

Table 18.1 (continued)

Gene	Race	Study design	Case	Control	Genotype distribution[a]	Comparison	Adjustment[b]	OR (95% CI)	Reference
MGMT L84F	Chinese	Population	162	251	76,22,1 86,13,1	LF vs. LL FF va. LL	Age	2.0 (1.2–3.3) 3.4 (0.3–38.1)	*(12)*
hHR23B A249V	Mixed (90% Caucasians)	Hospital	494	470	29[c]	AA vs. AV/VV	Age,BPH, FH, S	1.1 (0.8–1.4)	*(17)*
NBS1E185Q	unknown	Family	121	200	33,52,15 44,40,16	EQ vs. EE QQ vs. EE	none	1.4 (0.7–2.8) 0.8 (0.4–1.6)	*(28)*
		Population	200	200	41,47,12 44,40,16	EQ vs. EE QQ vs. EE		1.6 (0.9–2.9) 1.2 (0.7–2.3)	
XPC A499V	Mixed (90% Caucasians)	Hospital	494	470	24[c]	AA/AV vs. VV	Age,BPH, FH, S	0.9(0.5–1.5)	*(17)*
XPC K939Q	Mixed (90% Caucasians)	Hospital	494	470	38[c]	KK/KQ vs. QQ	Age,BPH, FH, S	1.0 (0.7–1.4)	*(17)*
XPC K939Q	Japanese	Hospital	165	165	47,47,6 44.42.14	KK/KQ vs. QQ	Age	2.5 (1.1–5.5)	*(10)*
XPD D312N	Mixed	Siblings	637	480	44,45,12 44,48, 8	DD vs. DN/NN	None	1.6 (1.0–2.5)	*(13)*
	Caucasian	Siblings	572	437	40,47,13 41,50, 9	DD vs. DN/NN	None	1.6 (1.0–2.5)	
XPD D312N	Mixed (90% Caucasians)	Hospital	494	470	41[c]	DD/DN vs. NN	Age,BPH, FH, S	0.8 (0.5–1.2)	*(17)*
XPD K751Q	Mixed (90% Caucasians)	Hospital	494	470	35[c]	KQ/QQ vs. KK	Age,BPH, FH, S	0.9 (0.6–1.4)	*(17)*
XPD K751Q	Chinese	Population	162	251	88,12,0 86,13,1	KQ/QQ vs. KK	Age	0.8 (0.5–1.5)	*(12)*

XPD K751Q	Mixed	Siblings	637	480	40,47,13 41,47,12	QQ vs. KQ/KK	None	1.1 (0.7–1.8)	*(13)*
	Caucasian	Siblings	572	437	39,48,13 41,47,12	QQ vs. KQ/KK	None	1.2 (0.7–2.0)	
XPF/ERCC4 R415Q	Mixed (90% Caucasians)	Hospital	494	470	9[c]	RQ vs. RR	Age, BPH, S. FH	1.4 (1.0–2.0)	*(17)*
XPG/ERCC5 D1104H	Mixed (90% Caucasians)	Hospital	494	470	46[c]	DD/DH vs. HH	Age, BPH, S, FH	0.8 (0.5–1.5)	*(17)*
XRCC1 R194W	Mixed (90% Caucasians)	Population	76	182	88,12,0 84,15,1	RW/WW vs. RR	Age, race	0.7 (0.3–1.6)	*(15)*
XRCC1 R194W	Japanese	Hospital	165	165	42,48,13 52,38,10	RW/WW vs. RR	Age	1.5 (0.9–2.2)	*(10)*
XRCC1 R194W	Chinese	Hospital	207	235	50,41,9 39,50,11	RW/WW vs. RR	Age,S,AL FH	0.6 (1.1–2.5)	*(30)*
XRCC1 R280H	Mixed (90% Caucasians)	Population	76	182	87,13,0 90,10,0	RH vs. RR	Age, race	1.5 (0.7–3.6)	*(15)*
XRCC1 R280H	Chinese	Hospital	207	235	80,19,1 82,17,1	RR/RH vs. HH	Age,S,AL FH	1.1 (0.7–1.9)	*(30)*
XRCC1 R399Q	Caucasian	Hospital	228	219	42,46,12 50,40,10	RQ vs. RR QQ vs. RR	Age, S, FH	1.6 (1.1–2.5) 1.6 (0.8–3.1)	*(11)*
	African Americans	Population	124	116	73,24,3 73,24,3	RQ vs. RR QQ vs. RR	Age, S	1.2 (0.6–2.2) 1.5 (0.3–8.2)	
XRCC1 R399Q	Chinese	Hospital	207	235	52,41,7 65,31,4	RR/RQ vs. QQ	Age,S,AL FH	1.7 (1.1–2.5)	*(30)*
XRCC1 R399Q	Chinese	Population	162	251	55,34,11 54,41,5	RQ vs. RR QQ vs. RR	Age	0.8 (0.5–1.3) 2.2 (1.0–4.8)	*(12)*
XRCC1 R399Q	Japanese	Hospital	165	165	53,38,9 52,42,6	RQ/QQ vs. RR	Age	1.0 (0.6–1.5)	*(10)*

(continued)

Table 18.1 (continued)

Gene	Race	Study design	Case	Control	Genotype distribution[a]	Comparison	Adjustment[b]	OR (95% CI)	Reference
XRCC1 R399Q	Mixed	Siblings	637	480	46,43,11 45,43,12	QQ vs. RQ/RR	None	0.9(0.6–1.4)	*(13)*
	Caucasians	Siblings	572	437	43,45,12 41,46,13	QQ vs. RQ/RR	None	0.9(0.5–1.4)	
XRCC1 R399Q	Mixed (90% Caucasians)	Population	76	182	49,39,12 42,43,15	RQ vs. RR QQ vs. RR	Age, race	0.8 (0.5–1.4) 0.7 (0.3–1.6)	*(15)*
XRCC1 R399Q	Mixed (92% Caucasians)	Population	77	174	49, (51)[d] 43, (57)	RQ vs. RR QQ vs. RR	none	0.8 (0.4–1.5) 0.7 (0.3–1.7)	*(26)*
XRCC3 T241M	Chinese	Population	162	251	87,11,2 87,13,1	TM vs. TT MM vs. TT	Age	0.8 (0.5–1.6) 2.2 (0.4–13.7)	*(12)*
XRCC7	Japanese	Hospital	165	165	7,48,45	gt/tt vs. gg	Age	1.0 (0.5–1.2)	*(10)*
g6721t					7,41,52				

[a]Numbers are percentages of each genotypes, and single number indicates percentage of minor allele frequency.[b]BPH benign prostate hyperplasia, FH: family history of prostate cancer, S, Smoking, AL, alcohol.[c]Minor allelic frequency.[d]Heteozygous and homozygous polymorphic genotypes were combined.

Table 18. 2
SNPs in DNA repair genes and their functional relation

Path-way	Gene	SNP ID no.	Minor allele frequency			Function	Reference
			White	Black	Asian		
BER	ADPRT V762A	rs1136410	0.11	0.06	0.33	Decrease enzyme activities in response to H_2O_2	*(27)*
BER	APE1 D148E	rs3136820	0.49	0.32	0.28	Hypersensitivity to ionizing radiation	*(65)*
BER	APE1 Q51H	rs1048945	0.03	0	0	Regulate the DNA binding activity	*(72)*
DRCC	ATM P1054R	rs1800057	0.02	0	0	Affect the cellular response after exposure to ionizing radiation	*(29)*
BER	hOGG1 S326C	rs1052133.	0.23	0.38	0.10	No difference in DNA repair capacity between genotypes 326Cys allele was associated with a decrease in p53 mutations Suppression of mutagenesis is lower in hOGG1 326Cys No difference in adducts between genotypes Adduct level was higher in CC genotypes	*(40–45,102)* *(46)* *(37)* *(39)* *(47,103.104)*
DR	MGMT I143V	rs2308321	0.13	0	0	More resistant to inactivation by MGMT pseudosubstrate, O^6-(4-bromothenyl) guanine	*(97)*
DR	MGMT L84F	rs12917	0.11	0.17	0.11	No affect on cell survival after exposure to N-methyl-N-nitro-N-nitrosoguanidine	*(96)*
NER	hHR23B A249V	rs1805329	0.15	0.26	0	No affect NER activity by plasmid-based NER assay	*(17)*
NER	X P C A499V	rs2228000	0.10	0.05	0	Affect NER activity by plasmid-based NER assay	*(17)*
NER	X P C K939Q	rs2228001	0.13	0.13	0	Linkage disequilibrium with intron 9, 5bp deletion which cause alternative splicing	*(105)*

(continued)

Table 18. 2
SNPs in DNA repair genes and their functional relation

Pathway	Gene	SNP ID no.	Minor allele frequency			Function	Reference
			White	Black	Asian		
NER	XPD/ERCC2 D312N	rs1799793	0.31	0	0.06	Affect NER activity by plasmid-based NER assay	*(17)*
NER	XPD/ERCC2 K751Q	rs1052559	0.27	0.04	0.17	Higher number of chromatid aberrations No association in SCE frequency or DNA adduct level Higher adduct level among never smokers	*(83)* *(84)* *(64)*
NER	XPF/ERCC4 R415Q	rs1800067	0.05	0	0	Reduced repair of aromatic DNA adducts	*(82)*
NER	XPG/ERCC5 D1104H	rs17655	0.27	0.44	0.46	Affect NER activity by plasmid-based NER assay	*(17)*
BER	XRCC1 R194W	rs1799782	0.05	0.24	0.08	No association with DNA-adduct levels, mutation rates, or sensitivity to ionizing radiation Lower bleomycin and benzo(a)pyrene diol epoxide sensitivity in vitro	*(65–67,69)* *(63)*
BER	XRCC1 R280H	rs25489	0.03	0.08	0.06	Higher bleomycin sensitivity	*(70,71)*
BER	XRCC1 R399Q	rs25487	0.47	0.45	0.10	Higher levels of aflatoxin B1-DNA adducts and higher bleomycin sensitivity	*(67,69,106)*
DSBR	XRCC3 T241M	rs861539	0.42	0.15	0.24	Hypersensitive to DNA-damaging agents	*(88)*

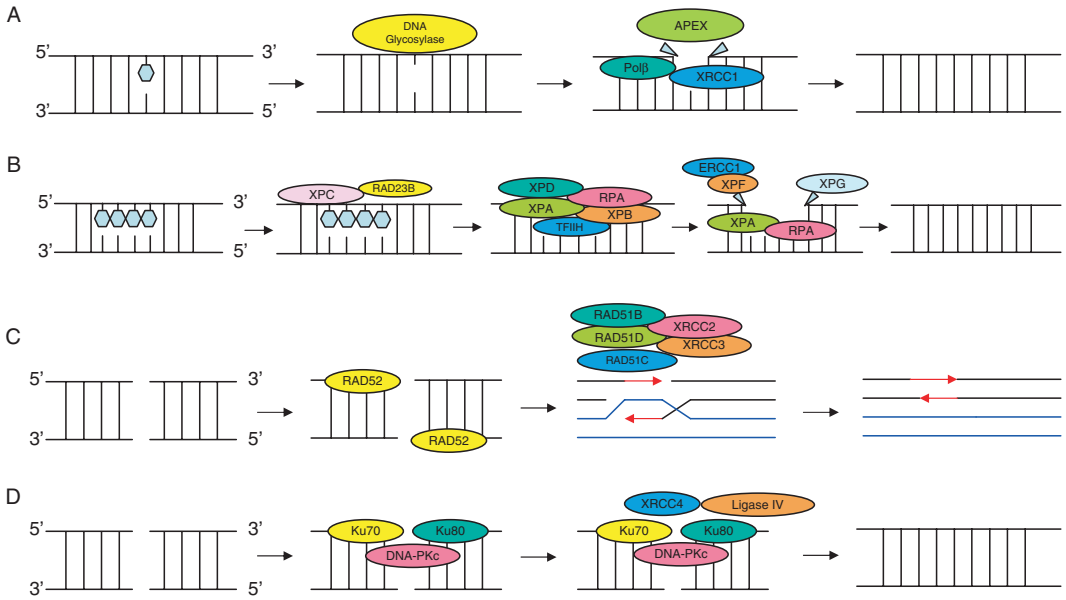


Fig. 18.1. (A) BER pathway targets DNA damaged during replication or by environmental agents. The single damaged base in DNA caused by endogenous metabolism or environmental oxidizing agents result in DNA adducts. BER involves removing the mutated base out of the DNA and repairing the base alone. (B) NER is associated with the repair of bulky adducts induced by several suspected environmental prostate cancer carcinogens. The NER pathway is a complex biochemical process that requires at least four steps: 1) damage recognition by a complex of bound proteins including XPC, XPA, and RPA; 2) unwinding of the DNA by TFIIH complex that includes XPD(ERCC2); 3) removal of the damaged single-stranded fragment (usually about 27–30 bp) by molecules including an ERCC1 and XPF complex and XPG; and 4) synthesis by DNA polymerases. (C) Double-strand breaks are produced by replication failure or by DNA-damaging agents. Two repair pathways exist to repair double strand breaks. The homologous recombination repair relies on DNA sequence complementarity between the intact chromatid and the damaged chromatid as the bases of stand exchange and repair. (D) The nonhomologous end-joining repair pathway requires direct DNA joining of the two double-strand-break ends.

polymerase b, and the nick is sealed by the DNA ligase III/XRCC1 complex, which acts as a scaffold for interaction with other BER enzymes *(36)*. It is expressed as 12 alternatively spliced isoforms with only the 1α-form containing a nuclear translocation signal *(37)*. Relatively high levels of expression of hOGG1 have been shown in several human tissues, including prostate *(14,20)*. Public database (http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?locusId=4968) lists 10 SNPs at the *hOGG1* locus *(38)*. *hOGG1* codon 326 polymorphism (rs1052123) in the 1α-specific exon 7 of the *hOGG1* results in an amino acid substitution from serine to cysteine **(Table 18.1)**. Results of studies for functional impact of the *hOGG1* S326C polymorphism are inconsistent **(Table 18.2)**. These studies used different measuring methods, high-performance liquid chromatography, flow cytometry, and different specimens, such as cell lines, leukocytes, and tissues. No difference in catalytic activities was observed between the *hOGG1* 326C and *hOGG1* 326S alleles in several studies *(39–45)*. However, the *hOGG1* encoded by the wild-type

326S allele exhibited higher DNA repair activity than the *hOGG1* 326C variant in other studies *(37,46–49)*.

The role of *hOGG1* 326 polymorphism in susceptibility to prostate cancer was assessed in four studies conducted in the United States and Canada *(14,16,18,25)*. The first was a population and family-based study that identified a significantly *decreased* risk associated with the *hOGG1* 326CC genotype *(16)*. This association was significant in nonfamilial prostate cancer patients, but not for familial prostate cancer. In contrast, the second, hospital-based study observed a positive relationship with prostate cancer risk *(14)*. The other two larger studies (996 and 439 cases) found no association between *hOGG1* S326C polymorphism and prostate cancer risk *(18,25)*. These inconsistent results could be explained by small sample sizes of first two studies (n = 84 and 245 cases). Epidemiologic studies of the *hOGG1* S326C polymorphism with risk of other cancers show consistent evidence for an increased risk for esophageal *(50)*, lung *(51–58)*, nasopharyngeal *(59)*, upper aerodigestive tract *(59–61)*, and colon *(62)* cancers.

Xu et al. *(16)* investigated other *hOGG1* polymorphisms in addition to codon 326 polymorphism. Among nine polymorphisms investigated in the *hOGG1* promoter region, significantly increased risks were observed for homozygous polymorphic genotypes of two variants compared with wild types (*hOGG1* a7143g and a11657g). A larger study (n = 996 cases) did not replicate the *hOGG1* a11657g finding *(25)*. The difference of these two studies is sample size. The first study observed a significant result based upon 1 control and 11 patients with homozygous *hOGG1* 11657gg genotype.

3.1.2. X-Ray Repair Cross-Complementing Group 1 (XRCC1)

After base excision by hOGG1 enzyme, the XRCC1/DNA ligase III complex seals the space *(36)*. Although 32 SNPs in *XRCC1* have been reported *(38)*, only three SNPs have been investigated as potential prostate cancer risk factors (R194W [rs17997820], R280H [rs25489], and R399Q [rs25487]). The functional significance of the *XRCC1* 194W allele is somewhat controversial. One study reported lower bleomycin and benzo(a)pyrene diol epoxide sensitivity in vitro *(63)*, but these results were not confirmed by other investigators *(64–67)* (**Table 18.2**). However, the *XRCC1 R*194W polymorphism may have detectable effects on DNA-adduct levels, mutation rates, or sensitivity to ionizing radiation *(65,66,68,69)*. The functional significance of codon R280H polymorphism is not yet well established; however, the codon 280 amino acid is located in the proliferating cell nuclear antigen binding region that has been associated with higher bleomycin sensitivity *(70,71)*. The *XRCC1* 399Q allele has been associated with higher levels of aflatoxin B1-DNA adducts and higher bleomycin sensitivity *(63,68,69)*.

Consistent with the functional data, the *XRCC1* R194W does not seem to influence risk in two small studies (n = 76 and 165 cases) *(10,15)*. However, a recent study suggested 194W allele provides a protective effect *(30)*. The *XRCC1* R280H polymorphism that has been evaluated in two small studies precluding any conclusion on risk of prostate cancer *(15,30)*.

The *XRCC1* R399Q polymorphism has been the most frequently investigated of the BER genes. Recently, Chen et al. *(11)* reported a significant association between *XRCC1* R399Q polymorphism and prostate cancer risk among Caucasians, but not in African Americans. Two studies in China observed a significantly increased risk with the *XRCC1* R399*Q* polymorphism *(12,30)*. The largest study to date *(13)* and three smaller studies *(10,15,26)* did not replicate the positive association. However, two small studies reported a significantly higher risk among men with the *XRCC1* 399 QQ genotype and low vitamin E intake/antioxidants *(15,26)*.

In summary, the *XRCC1* R399Q polymorphism has been associated with risk in five of seven studies, but only among men with low antioxidant, vitamin E intake in two of five studies. Additional studies are needed to clarify these potential associations.

3.1.3. Apurinic/Apyrimidic (A/P) Endonuclease (APE1)

When BER enzymes initiate repair, the phosphodiester bond at the 5′side of the intact apurinic/apyrimidinic site is incised by APE1, which is the rate-limiting enzyme. Six polymorphisms in *APE1* have been reported, including relatively common alleles at codons Q51H (rs1048945) and D148E (rs3136820) *(38)*. Although the functional significance of *APE1* 51H allele has not been reported, it is conserved in most mammals and located in the *Ref1* domain, which is essential for redox regulation of DNA binding proteins, such as p53 *(72)*. Therefore, *APE1* Q51H polymorphism may affect the ability of APE1 to regulate DNA binding activity. The *APE1* D148E polymorphism is associated with mitotic delay of lymphocytes from healthy subjects, implying higher sensitivity to ionizing radiation *(65)*. However, this variant was predicted as no impact on endonuclease and DNA binding activity in in vitro functional analysis *(72,73)*.

Both *APE1* Q51H and D148E polymorphisms have been examined in a hospital based study of Caucasians and African Americans. No associations between these polymorphisms and prostate cancer risk were observed *(11)*.

3.1.4. Poly(ADP-ribose) Polymerase (ADPRT)

ADPRT is involved in DNA-damage signaling, genomic stability of damaged cells, BER, recombination, and the transcriptional regulation of tumor suppressor genes *(74,75)*. ADPRT recognizes and binds DNA damage, recruits other DNA-repair enzymes to the site of damage, and provides support for ligation *(76)*. Twenty-five polymorphisms in *ADPRT* have been

reported, including the relatively common *ADPRT* V762A (rs1136410) *(38)*.

The change from valine to alanine moves the codon 762 residue further away from the codon 888G residue, which is a part of the active site *(76)*. Locket et al. *(27)* observed that *ADPRT* V762A is significantly associated with prostate cancer risk and decreased enzyme function in response to oxidative damage.

3.2. NER Pathway

NER is associated with the repair of bulky adducts *(77,78)* induced by several suspected environmental prostate cancer carcinogens, such as polycyclic aromatic hydrocarbons, heterocyclic aromatic amines from well-done meats, and pesticides. The NER pathway is a complex biochemical process that requires 20–25 enzymes and at least four steps: 1) damage recognition by a complex of bound proteins, including xeroderma pigmentosum complementation group C (XPC*),* XPA, and replication protein A (RPA); 2) unwinding of the DNA by the transcription factor IIH (TFIIH) complex that includes XPD(ERCC2); 3) removal of the damaged single-stranded fragment (usually about 27–30 bp) by molecules including an ERCC1 and XPF complex and XPG; and 4) synthesis by DNA polymerases **(*(6)*; Fig. 18.1B)**.

3.2.1. Xeroderma Pigmentosum Complementation Group D (XPD)

The *XPD* (*ERCC2*) gene product is a subunit of TFIIH (DNA helicase), promotes bubble formation, and is necessary for NER and transcription. Fourteen polymorphisms in *XPD* have been reported *(38)*, including common alleles at codons D312N (rs1799793) and K751Q (rs1052559). Several studies reported that subjects with wild-type genotypes for *XPD* K751Q and D312N polymorphisms exhibit the highest NER activity, whereas homozygous variant genotypes of either polymorphism show low NER activity *(79,80)*. Hou et al. *(81,82)* reported that the *XPD* 312N allele has reduced capacity to repair aromatic DNA adducts. Lunn et al. *(83)* reported that *XPD* K751Q was associated with higher levels of chromatid aberrations in white blood cells. Conversely, Duell et al. *(84)* evaluated phenotypic effects of codons 312 and 751 polymorphisms by measuring two markers of DNA damage, sister chromatid exchange (SCE) frequencies, and polyphenol DNA adducts. Both polymorphisms were unrelated to SCE frequency or DNA adduct level *(84)*.

A potential role of *XPD* codons D312N and K751Q with prostate cancer risk has been investigated in three studies *(12,13,17)*. All three studies observed no association between *XPD* K751Q polymorphism and prostate cancer risk in U.S. and Chinese populations. Although Rybicki et al. *(13)* observed a significant risk increase with the D312N polymorphism, this was not replicated by Lockett et al. *(17)* after adjustment for age, BPH, family history, and smoking.

3.2.2. Xeroderma Pigmentosum Complementation Group F (XPF)/(ERCC4)

XPF is a key enzyme responsible for excising bulky adducts from damaged DNA. XPF interacts with ERCC1 to form a complex that is required to repair DNA interstrand cross-linking damage *(85)*. Ten polymorphisms in *XPF* have been reported *(38)*, but only R415Q (rs1800067) has been studied in an epidemiologic investigation. Lockett et al. *(17)* reported that *XPF* R415Q polymorphism was associated with a moderate, near significant increase in prostate cancer risk (odds ratio [OR] = 1.4).

3.2.3. Xeroderma Pigmentosum Complementation Group G (XPG)/(ERCC5)

XPG is responsible for a structure-specific endonuclease activity that is essential for the two incision steps in NER *(85)*. The XPG enzyme has been suggested to act on the single-stranded region created as a result of the combined action of the XPB helicase and the XPD helicase at the DNA damage site. XPG incises the 3′ side of damaged DNA before the 5′ incision made by XPF-ERCC1 complexes. XPG has a structural function in the complex of the DNA-hR23B. Twelve SNPs were reported including *XPG* D1104H (rs17655) *(38)*. The functional effects of D1104H SNP are still unknown, but the lack of association with prostate cancer risk *(17)* decreases the incentives to pursue small studies.

3.2.4. Xeroderma Pigmentosum Complementation Group C (XPC)

In the early steps of the NER process, the XPC-hR23B protein complex has a structure-specific affinity for certain defined lesions. Thus, this complex can bind damaged DNA and change the DNA conformation around the lesion. DNA damage recognition is carried out by the XPC-hR23B protein complex *(86)*, followed by recruitment of the TFIIH complex. Among twenty known SNPs *(38)*, two common polymorphisms at codons A499V (rs2228000) and K939Q (rs2228001) have been investigated. There are no published data on their potential functional significance. Lockett et al. *(17)* observed no significant association between these polymorphisms and prostate cancer risk. However, a small study (n = 165 cases) observed a significant 2.5-fold risk increase among Japanese men with the 939K allele *(10)*.

3.2.5. Human Homolog RAD23B (hR23B)

The hR23B enzyme, which is the human homolog of the yeast NER protein RAD23, forms a complex with XPC. The XPC-hR23B-TFIIH complex unwinds the DNA duplex around the damaged site. Five SNPs have been reported *(38)*, but only the A249V has been investigated. Lockett et al. *(17)* found no association between this SNP and prostate cancer risk, and the absence of functional data tempers interest.

3.3. DSBR Pathway

Double-strand breaks are produced by replication failure or by DNA-damaging agents such as ionizing radiation. Two repair pathways exist to repair double strand breaks: 1) the homologous recombination repair relies on DNA sequence complementarity

between the intact chromatid and the damaged chromatid as the bases of stand exchange and repair (Fig. 1C); and 2) the nonhomologous end-joining repair pathway requires direct DNA joining of the two double-strand-break ends (*[87]*; **Fig. 18.1D**).

3.3.1. Xeroderma Pigmentosum Complementation Group 3 (XRCC3)

XRCC3 is involved in homologous recombination repair process, and at least six SNPs have been identified *(38)*. Araujo et al. *(88)* reported that the variant XRCC3 enzyme (T241M) was functionally active for homology-directed repair (HDR) determined by a quantitative fluorescence assay. They also found that cells expressing this variant have been found to be no more sensitive to DNA-damaging agents than cells expressing the wild-type enzyme *(88)*.

XRCC3 T241M polymorphism has been analyzed in relation to prostate cancer risk in a population-based study in China *(12)*. This relatively small study detected no statistically significant association between *XRCC3* T241M polymorphism and prostate cancer risk, but homozygous carriers deserve further study. Relative to men with the TT genotype and a low intake of preserved foods, those with the MT+MM genotype and a higher intake of nitrosamines and nitrosamine precursors, had a significantly higher risk of prostate cancer (OR = 2.6; 95% confidence interval [CI] = 1.1–6.1). In contrast, men with the MT+MM genotype and a low intake of preserved foods had a significant reduction in risk (OR = 0.3; 95% CI = 0.1–0.96). These data suggest that diet factors, such as preserved foods, may influence prostate cancer risk in combination with genetic susceptibility in DNA repair pathways.

3.3.2. Xeroderma Pigmentosum Complementation Group 7 (XRCC7)

XRCC7/PRKDC (protein kinase, DNA-activated, catalytic polypeptide) is a key enzyme that becomes activated upon incubation with DNA. Genetic defects in this enzyme result in immunodeficiency, radiosensitivity, and premature aging *(89,90)*. These phenotypes are due to the defect of DNA double-strand break repair processes. Recent studies reveal that XRCC7 also participates in signal transduction cascades related to apoptotic cell death, telomere maintenance and other pathways of genome surveillance *(91)*. Only one epidemiologic study has been reported, and only one of the nine known SNPs *(38)*, the g6721t polymorphism located intron 8, was investigated. No significant association between *XRCC7* g6721t polymorphism and prostate cancer risk was observed in this small hospital-based study of Japanese men *(10)*. The functional significance of the *XRCC7* g6721t polymorphism is not firmly established, but it may regulate splicing and cause mRNA instability *(92)*.

3.3.3. Nijmegen Breakage Syndrome1 (NBS1)

The Nijmegen breakage syndrome 1 (*NBS1*) is part of a protein complex that forms in response to DNA damage to maintain chromosomal integrity. The exact role of NBS1 in DNA repair is

not fully understood because NBS1 does not have a DNA binding site or kinase activity, which is usually required in DNA repair. However, the N-terminal domain binds to γH2AX, and this may be an important step to recruit the NBS1 protein complex to the proximity of DNA repair *(93)*. Thirty-eight polymorphisms in *NBS1* have been reported, including codon E185Q (rs1805794) *(38)*. Although there is no information regarding changes in the activity of the *NBS1*-185Q variant, the region between amino acid 108–196 of the NBS1 enzyme constitutes a *BRCA1* COOH terminus domain that is presumably involved in cell-cycle checkpoints or in DNA repair *(94)*. In this same report, all individuals with the *NBS1* 185QQ genotype had lung tumors with *p53* mutations in contrast with only 46% of *p53* mutations in tumors from individuals with 185EE genotype *(94)*. In the only study of this variant in relation to prostate cancer, Hebbring et al. *(28)* observed that *NBS1* E185Q polymorphism was not strongly associated with familial or sporadic prostate cancer risk.

3.4. Direct Reversal (DR) Pathway

The biologically significant DNA lesions produced by both carcinogenic and chemotherapeutic alkylating agents are O^6-alkylguanine adducts, which can pair with thymine instead of cytosine during DNA replication. Therefore, O^6-alkylguanine adducts may be responsible for the increase in the frequency of mutations following exposure to alkylating agents, and carcinogenesis *(95)*.

3.4.1. Methylguanine-DNA Methyltransferase (MGMT)

The only known enzyme in the DR pathway is MGMT. MGMT transfers the alkyl group at the O^6 position of guanine to a cysteine residue within its active site, leading to the direct restoration of the natural chemical composition of DNA without the need for genomic reconstruction. Defective MGMT activity often increases mutation because O^6-MeG mispairs with thymine during DNA replication *(87)*.

Among 16 SNPs in *MGMT (38)*, the functional effects of two common SNPs (L84F and I143V) have been examined *(12)*. Although L84F polymorphism does not affect cell survival after exposure to N-methyl-N-nitro-N-nitrosoguanidine *(96)*, the *MGMT* 143V allele is significantly more resistant to inactivation by MGMT pseudosubstrate, O^6-(4-bromothenyl)guanine *(97)*. However, Liu et al. *(95)* reported that the relative gene expression level, evaluated by the real-time reverse transcription-polymerase chain reaction assay of MGMT in peripheral lymphocytes, is not significantly different between prostate cancer patients and age- and ethnicity-matched controls *(20)*. Furthermore, this I143V change may affect the isoleucine residue close to the alkyl acceptor cysteine residue at codon 145 *(95)*.

Ritchey et al. *(12)* examined *MGMT* L84F and I143V polymorphisms in a population-based case-control study of Chinese (162 cases, 251 controls). The *MGMT* L84F polymorphism was

significantly associated with a two-fold increased risk, but the I143V polymorphism was not.

3.5. Damage Recognition Cell Cycle Delay Responses

Minimizing transmission of DNA mutations to daughter cells is biologically important. Therefore, some enzymes can recognize DNA damage and signal the status to initiate DNA replication *(87)*. DNA damage activates a cell cycle delay response pathway to earn time for damage repair *(98)*. Defects in this pathway may result in genomic instability, ultimately leading to cancer susceptibility. The key enzyme of this damage recognition cell delay response pathway is the ataxia telangiectasia-mutated (ATM) and the tumor suppressor protein p53.

3.5.1. ATM Protein

ATM, which is the product of the gene mutated in patients with the autosomal recessive disorder ataxia telangiectasia, is a key enzyme responsible for downstream signaling. ATM is activated by DNA damage and induces the transactivation of various proteins involved in cell cycle arrest, apoptosis, DNA repair, and centrosome duplication. In particular, ATM regulates phosphorylation of p53 protein, thereby allowing p53 to accumulate. ATM also regulates a variety of downstream proteins, including the tumor suppressor BRCA1, checkpoint kinase CHK2, checkpoint protein RAD50, and DNA repair protein *NBS1 (99)*. Nine polymorphisms in ATM have been reported *(38)*. Angele et al. *(29)* investigated the association of five SNPs in ATM (D1853N, D1853V, ivs38-8t>c, ivs38-15g>c, and P1054R) with prostate cancer risk. The ATM P1054R variant is located in the β-adaptin domain of the ATM protein and has been linked to an increased cancer risk, particularly breast cancer *(100,101)*. Only the ATM 1054R allele was significantly associated with an increased risk of prostate cancer *(29)*. Furthermore, in the same study, a lymphoblastoid cell line carrying P1054R polymorphism shows a significantly different cell progression to that seen in cell lines that carry a wild type ATM after exposure to ionizing radiation. These results suggest that the codon 1054 polymorphism confers an altered cellular phenotype and might be associated with prostate cancer risk.

3.6. Oligogenic Model

Results of epidemiologic studies have been inconsistent. Although the exact basis for the inconsistency is unknown, a number of factors may be relevant, including various study design limitations (e.g., using mixed ethnic groups, polymorphisms with unknown functional effects, enzymes not expressed in target tissues, and use of prevalent cases), competing or overlapping DNA repair pathways, and grouping of genotypes, small sample sizes, or variations in allelic frequencies across populations. Many of the studies used convenience samples of cases and controls. However, perhaps the main reason is investigating only one SNP and one gene from a complex biological pathway.

Due to recent advance in high-throughput genotyping techniques, multiple polymorphisms within genes, multiple genes in the same pathway, and haplotype approaches are now available to greatly increase the depth of exploration. Although several studies analyze multiple SNPs within a gene, only two studies used a haplotype analysis *(10,11)*. A few studies also analyzed multiple genes in the DNA repair pathway. This approach may provide more biologically plausible insight into the studied associations, including interaction effects of different alleles on prostate cancer risk.

When prostate cancer risk for combined effects of multiple polymorphisms in different DNA repair genes were estimated, we often find much stronger associations. Rybicki et al. *(13)* reported that the OR for the combined effects of the *XPD* 312 DD and *XRCC1* 399 QQ genotype was 4.8 compared with *XPD* 312 DN/NN and *XRCC1* RR/RQ genotypes. In a separate study, similar combined effects were observed in individuals with *APE1* D148E/*XRCC1* R399Q polymorphisms. The OR for the combined effects of the *APE1* 51QQ and *XRCC1* 399RQ/QQ genotypes was 4.0 compared with *APE1* QH/HH and *XRCC1* 399RR *(11)*. Recently, Hirata et al. *(10)* reported that significant combined effects of SNPs in *XPC* and *XRCC1* when two genes from different DNA repair pathway, were observed.

These combined effects of multiple SNPs and different genes suggest that severely defected DNA repair capacity may play a role in prostate cancer risk, particularly when the function of multiple DNA repair genes are compromised.

4. Discussion

Fifteen published epidemiologic studies have explored the potential association of 31 SNPs in 14 DNA repair genes with prostate cancer risk. Although more studies are warranted, the only pathway that shows significant associations is BER. The XRCC*1* 399Q allele is associated with increased risk for carriers alone or when the variant allele is combined with other DNA repair polymorphisms or low antioxidant diet *(10–13,15,26)*. Lockett et al. *(27)* reported that *ADPRT* V762A variant contributed to prostate cancer risk and altered enzyme activity. The *hOGG1* S326C polymorphism needs additional studies. Particularly, results from epidemiologic studies of other cancer sites show a consistent relation with increased risk *(50–62)*.

SNPs in two NER genes, *XPC* and *XPD*, show significant associations with prostate cancer risk in some *(10,13)*, but not all studies *(17)*. Finally, a study of ATM yielded promising results that merit additional research *(29)*.

Epidemiologic studies of SNPs in DNA repair genes may inform individual susceptibility and provide insight on potential mechanisms of carcinogenesis. The current challenge is to validate the functional impact of important SNPs identified by epidemiological studies. Another challenge is to identify "causal SNPs" through epidemiologic studies, especially in studies investigating the role of SNPs in disease as complex as prostate cancer. Results of many epidemiologic studies lack statistical significance. Most studies do not have enough power to investigate gene–gene and gene–environmental interactions. Studies investigating a single SNP in a DNA repair gene are not likely detecting difference in overall DNA repair activity. As we presented in the oligogenic model section, large studies investigating multi-SNPs and multigenes may generate significant data through combined genotype and haplotype analysis.

In the future, relatively inexpensive high-throughput genotyping methods and more functional data will be available based on an individual's genetic profile that affects the progression, metastasis, and response to therapy. The interpretation of epidemiological data and translation to patient care will be accelerated through pooled analysis and consortia.

References

1. American Cancer Society. (2007). Cancer Facts & Figures 2007. American Cancer Society, Atlanta, GA.
2. Hsing, A.W., L. Tsao, and S.S. Devesa. (2000). International trends and patterns of prostate cancer incidence and mortality. Int J Cancer **85**, 60–7.
3. Crawford, E.D. (2003) Epidemiology of prostate cancer. Urology **62**, 3–12.
4. Noble, R.L. (1977). The development of prostatic adenocarcinoma in Nb rats following prolonged sex hormone administration. Cancer Res **37**, 1929–33.
5. Henderson, B.E., R.K. Ross, M.C. Pike, and J.T. Casagrande (1982). Endogenous hormones as a major factor in human cancer. Cancer Res **42**, 3232–9.
6. Friedberg, E.C. (2001). How nucleotide excision repair protects against cancer. Nat Rev Cancer **1, 22**–33.
7. Mullaart, E., P.H. Lohman, F. Berends, and J. Vijg. (1990). DNA damage metabolism and aging. Mutat Res **237**, 189–210.
8. Wood, R.D., M. Mitchell, J. Sgouros, and T. Lindahl. (2001). Human DNA repair genes. Science **291**, 1284–9.
9. Goode, E.L., C.M. Ulrich, and J.D. Potter. (2002). Polymorphisms in DNA repair genes and associations with cancer risk. Cancer Epidemiol Biomarkers Prev **11**, 1513–30.
10. Hirata, H., Y. Hinoda, Y. Tanaka, N. Okayama, Y. Suehiro, K. Kawamoto, N. Kikuno, S. Majid, K. Vejdani, and R. Dahiya. (2007). Polymorphisms of DNA repair genes are risk factors for prostate cancer. Eur J Cancer **43**, 231–7.
11. Chen, L., C.B. Ambrosone, J. Lee, T.A. Sellers, J. Pow-Sang, and J.Y. Park. (2006). Association between polymorphisms in the DNA repair genes XRCC1 and APE1, and the risk of prostate cancer in white and black Americans. J Urol. **175**, 108–12; discussion 112.
12. Ritchey, J.D., W.Y. Huang, A.P. Chokkalingam, Y.T. Gao, J. Deng, P. Levine, F.Z. Stanczyk, and A.W. Hsing. (2005). Genetic variants of DNA repair genes and prostate cancer: a population-based study. Cancer Epidemiol Biomarkers Prev **14**, 1703–9.
13. Rybicki, B.A., D.V. Conti, A. Moreira, M. Cicek, G. Casey, and J.S. Witte. (2004). DNA repair gene XRCC1 and XPD polymorphisms and risk of prostate cancer. Cancer Epidemiol Biomarkers Prev **13**, 23–9.
14. Chen, L., A. Elahi, J. Pow-Sang, P. Lazarus, and J. Park. (2003). Association between polymorphism of human oxoguanine glyco-

sylase 1 and risk of prostate cancer. J Urol. **170**, 2471–4.

15. van Gils, C.H., R.M. Bostick, M.C. Stern, and J.A. Taylor. (2002). Differences in base excision repair capacity may modulate the effect of dietary antioxidant intake on prostate cancer risk: an example of polymorphisms in the XRCC1 gene. Cancer Epidemiol Biomarkers Prev **11**, 1279–84.
16. Xu, J., S.L. Zheng, A. Turner, S.D. Isaacs, K.E. Wiley, G.A. Hawkins, B.L. Chang, E.R. Bleecker, P.C. Walsh, D.A. Meyers, and W.B. Isaacs. (2002). Associations between hOGG1 sequence variants and prostate cancer susceptibility. Cancer Res **62**, 2253–7.
17. Lockett, K.L., I.V. Snowhite, and J.J. Hu. (2005). Nucleotide-excision repair and prostate cancer risk. Cancer Lett **220**, 125–35.
18. Nock, N.L., M.S. Cicek, L. Li, X. Liu, B.A. Rybicki, A. Moreira, S.J. Plummer, G. Casey, and J.S. Witte. (2006). Polymorphisms in estrogen bioactivation, detoxification and oxidative DNA base excision repair genes and prostate cancer risk. Carcinogenesis. **27**, 1842–8.
19. Xu, Z., L.X. Qian, L.X. Hua, X.R. Wang, J. Yang, W. Zhang, and H.F. Wu. (2007). Relationship between DNA repair gene XRCC1 Arg399Gln polymorphism and susceptibility to prostate cancer in the Han population in Jiangsu and Anhui. Zhonghua Nan Ke Xue **13**, 327–31.
20. Liu, Z., L.E. Wang, S.S. Strom, M.R. Spitz, R.J. Babaian, J. DiGiovanni, and Q. Wei. (2003). Overexpression of hMTH in peripheral lymphocytes and risk of prostate cancer: a case-control analysis. Mol Carcinog **36**, 123–9.
21. Strom, S.S., M.R. Spitz, Y. Yamamura, R.J. Babaian, P.T. Scardino, and Q. Wei (2001). Reduced expression of hMSH2 and hMLH1 and risk of prostate cancer: a case-control study. Prostate **47**, 269–75.
22. Chen, Y., J. Wang, M.M. Fraig, K. Henderson, N.K. Bissada, D.K. Watson, and C.W. Schweinfest. (2003). Alterations in PMS2, MSH2 and MLH1 expression in human prostate cancer. Int J Oncol **22**, 1033–43.
23. Chen, Y., J. Wang, M.M. Fraig, J. Metcalf, W.R. Turner, N.K. Bissada, D.K. Watson, and C.W. Schweinfest. (2001). Defects of DNA mismatch repair in human prostate cancer. Cancer Res **61**, 4112–21.
24. Hu, J.J., M.C. Hall, L. Grossman, M. Hedayati, D.L. McCullough, K. Lohman, and L.D. Case. (2004). Deficient nucleotide excision repair capacity enhances human prostate cancer risk. Cancer Res **64**, 1197–201.
25. Nam, R.K., W.W. Zhang, M.A. Jewett, J. Trachtenberg, L.H. Klotz, M. Emami, L. Sugar, J. Sweet, A. Toi, and S.A. Narod. (2005). The use of genetic markers to determine risk for prostate cancer at prostate biopsy. Clin Cancer Res **11**, 8391–7.
26. Goodman, M., R.M. Bostick, K.C. Ward, P.D. Terry, C.H. van Gils, J.A. Taylor, and J.S. Mandel. (2006). Lycopene intake and prostate cancer risk: effect modification by plasma antioxidants and the XRCC1 genotype. Nutr Cancer **55**, 13–20.
27. Lockett, K.L., M.C. Hall, J. Xu, S.L. Zheng, M. Berwick, S.C. Chuang, P.E. Clark, S.D. Cramer, K. Lohman, and J.J. Hu. (2004). The ADPRT V762A genetic variant contributes to prostate cancer susceptibility and deficient enzyme function. Cancer Res **64**, 6344–8.
28. Hebbring, S.J., H. Fredriksson, K.A. White, C. Maier, C. Ewing, S.K. McDonnell, S.J. Jacobsen, J. Cerhan, D.J. Schaid, T. Ikonen, V. Autio, T.L. Tammela, K. Herkommer, T. Paiss, W. Vogel, M. Gielzak, J. Sauvageot, J. Schleutker, K.A. Cooney, W. Isaacs, and S.N. Thibodeau. (2006). Role of the nijmegen breakage syndrome 1 gene in familial and sporadic prostate cancer. Cancer Epidemiol Biomarkers Prev **15**, 935–8.
29. Angele, S., A. Falconer, S.M. Edwards, T. Dork, M. Bremer, N. Moullan, B. Chapot, K. Muir, R. Houlston, A.R. Norman, S. Bullock, Q. Hope, J. Meitz, D. Dearnaley, A. Dowe, C. Southgate, A. Ardern-Jones, D.F. Easton, R.A. Eeles, and J. Hall. (2004). ATM polymorphisms as risk factors for prostate cancer development. Br J Cancer **91**, 783–7.
30. Xu, Z., L.X. Hua, L.X. Qian, J. Yang, X.R. Wang, W. Zhang, and H.F. Wu. (2007). Relationship between XRCC1 polymorphisms and susceptibility to prostate cancer in men from Han, Southern China. Asian J Androl **9**, 331–8.
31. Demple, B. and L. Harrison. (1994). Repair of oxidative damage to DNA: enzymology and biology. Annu Rev Biochem **63**, 915–48.
32. Robson, C.N., and I.D. Hickson. (1991). Isolation of cDNA clones encoding a human apurinic/apyrimidinic endonuclease that corrects DNA repair and mutagenesis defects in E. coli xth (exonuclease III) mutants. Nucleic Acids Res **19**, 5519–23.
33. Wilson, D.M., 3rd, and D. Barsky. (2001). The major human abasic endonuclease: formation, consequences and repair of abasic lesions in DNA. Mutat Res **485**, 283–307.
34. Boiteux, S. and J.P. Radicella. (2000). The human OGG1 gene: structure, functions,

and its implication in the process of carcinogenesis. Arch Biochem Biophys **377**, 1–8.

35. Sunaga, N., T. Kohno, K. Shinmura, T. Saitoh, T. Matsuda, R. Saito, and J. Yokota. (2001). OGG1 protein suppresses G:C-->T: A mutation in a shuttle vector containing 8-hydroxyguanine in human cells. Carcinogenesis **22**, 1355–62.
36. Lindahl, T., and R.D. Wood. (1999). Quality control by DNA repair. Science. **286**, 1897–905.
37. Kohno, T., K. Shinmura, M. Tosaka, M. Tani, S.R. Kim, H. Sugimura, T. Nohmi, H. Kasai, and J. Yokota. (1998). Genetic polymorphisms and alternative splicing of the hOGG1 gene, that is involved in the repair of 8-hydroxyguanine in damaged DNA. Oncogene **16**, 3219–25.
38. National Center for Biotechnology Information (2006). SNP500 Cancer. Cancer Genome Anatomy Project. Cited in http://snp500cancer.nci.nih.gov.
39. Shinmura, K., T. Kohno, H. Kasai, K. Koda, H. Sugimura, and J. Yokota. (1998). Infrequent mutations of the hOGG1 gene, that is involved in the excision of 8-hydroxyguanine in damaged DNA, in human gastric cancer. Jpn. J. Cancer Res. **89**, 825–8.
40. Janssen, K., K. Schlink, W. Gotte, B. Hippler, B. Kaina, and F. Oesch. (2001). DNA repair activity of 8-oxoguanine DNA glycosylase 1 (OGG1) in human lymphocytes is not dependent on genetic polymorphism Ser326/Cys326. Mutat Res **486**, 207–16.
41. Dherin, C., J.P. Radicella, M. Dizdaroglu, and S. Boiteux. (1999). Excision of oxidatively damaged DNA bases by the human alpha-hOgg1 protein and the polymorphic alpha-hOgg1(Ser326Cys) protein which is frequently found in human populations. Nucleic Acids Res **27**, 4001–7.
42. Hardie, L.J., J.A. Briggs, L.A. Davidson, J.M. Allan, R.F. King, G.I. Williams, and C.P. Wild. (2000). The effect of hOGG1 and glutathione peroxidase I genotypes and 3p chromosomal loss on 8-hydroxydeoxyguanosine levels in lung cancer. Carcinogenesis **21**, 167–72.
43. Park, Y.J., E.Y. Choi, J.Y. Choi, J.G. Park, H.J. You, and M.H. Chung. (2001). Genetic changes of hOGG1 and the activity of oh8Gua glycosylase in colon cancer. Eur J Cancer **37**, 340–6.
44. Kondo, S., S. Toyokuni, T. Tanaka, H. Hiai, H. Onodera, H. Kasai, and M. Imamura. (2000). Overexpression of the hOGG1 gene and high 8-hydroxy-2'-deoxyguanosine (8-OHdG) lyase activity in human colorectal carcinoma: regulation mechanism of the 8-OHdG level in DNA. Clin Cancer Res **6**, 1394–400.
45. Blons, H., J.P. Radicella, O. Laccourreye, D. Brasnu, P. Beaune, S. Boiteux, and P. Laurent-Puig. (1999). Frequent allelic loss at chromosome 3p distinct from genetic alterations of the 8-oxoguanine DNA glycosylase 1 gene in head and neck cancer. Mol Carcinog **26**, 254–60.
46. Hu, Y.C. and S.A. Ahrendt. (2005). hOGG1 Ser326Cys polymorphism and G:C-to-T: A mutations: no evidence for a role in tobacco-related non small cell lung cancer. Int J Cancer **114**, 387–93.
47. Tarng, D.C., T.J. Tsai, W.T. Chen, T.Y. Liu, and Y.H. Wei. (2001). Effect of human OGG1 1245C-->G gene polymorphism on 8-hydroxy-2'-deoxyguanosine levels of leukocyte DNA among patients undergoing chronic hemodialysis. J Am Soc Nephrol **12**, 2338–47.
48. Chen, S.K., W.A. Hsieh, M.H. Tsai, C.C. Chen, A.I. Hong, Y.H. Wei, and W.P. Chang. (2003). Age-associated decrease of oxidative repair enzymes, human 8-oxoguanine DNA glycosylases (hOgg1), in human aging. J Radiat Res **44**, 31–5.
49. Yamane, A., T. Kohno, K. Ito, N. Sunaga, K. Aoki, K. Yoshimura, H. Murakami, Y. Nojima, and J. Yokota. (2004). Differential ability of polymorphic OGG1 proteins to suppress mutagenesis induced by 8-hydroxyguanine in human cell in vivo. Carcinogenesis **25**, 1689–94.
50. Xing, D.Y., W. Tan, N. Song, and D.X. Lin. (2001). Ser326Cys polymorphism in hOGG1 gene and risk of esophageal cancer in a Chinese population. Int J Cancer **95**, 140–3.
51. Sugimura, H., T. Kohno, K. Wakai, K. Nagura, K. Genka, H. Igarashi, B.J. Morris, S. Baba, Y. Ohno, C. Gao, Z. Li, J. Wang, T. Takezaki, K. Tajima, T. Varga, T. Sawaguchi, J.K. Lum, J.J. Martinson, S. Tsugane, T. Iwamasa, K. Shinmura, and J. Yokota. (1999). hOGG1 Ser326Cys polymorphism and lung cancer susceptibility. Cancer Epidemiol Biomarkers Prev **8**, 669–74.
52. Wikman, H., A. Risch, F. Klimek, P. Schmezer, B. Spiegelhalder, H. Dienemann, K. Kayser, V. Schulz, P. Drings, and H. Bartsch (2000). hOGG1 polymorphism and loss of heterozygosity (LOH): significance for lung cancer susceptibility in a Caucasian population. Int J Cancer **88**, 932–7.
53. Ito, H., N. Hamajima, T. Takezaki, K. Matsuo, K. Tajima, S. Hatooka, T. Mitsudomi, M. Suyama, S. Sato, and R. Ueda. (2002). A limited association of OGG1 Ser326Cys

polymorphism for adenocarcinoma of the lung. J Epidemiol **12**, 258–65.

54. Le Marchand, L., T. Donlon, A. Lum-Jones, A. Seifried, and L.R. Wilkens. (2002). Association of the hOGG1 Ser326Cys polymorphism with lung cancer risk. Cancer Epidemiol Biomarkers Prev **11**, 409–12.
55. Sunaga, N., T. Kohno, N. Yanagitani, H. Sugimura, H. Kunitoh, T. Tamura, Y. Takei, S. Tsuchiya, R. Saito, and J. Yokota. (2002). Contribution of the NQO1 and GSTT1 polymorphisms to lung adenocarcinoma susceptibility. Cancer Epidemiol Biomarkers Prev **11**, 730–8.
56. Lan, Q., J.L. Mumford, M. Shen, D.M. Demarini, M.R. Bonner, X. He, M. Yeager, R. Welch, S. Chanock, L. Tian, R.S. Chapman, T. Zheng, P. Keohavong, N. Caporaso, and N. Rothman. (2004). Oxidative damage-related genes AKR1C3 and OGG1 modulate risks for lung cancer due to exposure to PAH-rich coal combustion emissions. Carcinogenesis **25**, 2177–81.
57. Park, J., L. Chen, M.S. Tockman, A. Elahi, and P. Lazarus. (2004). The human 8-oxoguanine DNA N-glycosylase 1 (hOGG1) DNA repair enzyme and its association with lung cancer risk. Pharmacogenetics **14**, 103–9.
58. Hung, R.J., P. Brennan, F. Canzian, N. Szeszenia-Dabrowska, D. Zaridze, J. Lissowska, P. Rudnai, E. Fabianova, D. Mates, L. Foretova, V. Janout, V. Bencko, A. Chabrier, S. Borel, J. Hall, and P. Boffetta. (2005). Large-scale investigation of base excision repair genetic polymorphisms and lung cancer risk in a multicenter study. J Natl Cancer Inst **97**, 567–76.
59. Cho, E.Y., A. Hildesheim, C.J. Chen, M.M. Hsu, I.H. Chen, B.F. Mittl, P.H. Levine, M.Y. Liu, J.Y. Chen, L.A. Brinton, Y.J. Cheng, and C.S. Yang. (2003). Nasopharyngeal carcinoma and genetic polymorphisms of DNA repair enzymes XRCC1 and hOGG1. Cancer Epidemiol Biomarkers Prev **12**, 1100–4.
60. Hao, B., H. Wang, K. Zhou, Y. Li, X. Chen, G. Zhou, Y. Zhu, X. Miao, W. Tan, Q. Wei, D. Lin, and F. He. (2004). Identification of genetic variants in base excision repair pathway and their associations with risk of esophageal squamous cell carcinoma. Cancer Res **64**, 4378–84.
61. Elahi, A., Z. Zheng, J. Park, K. Eyring, T. McCaffrey, and P. Lazarus. (2002). The human OGG1 DNA repair enzyme and its association with orolaryngeal cancer risk. Carcinogenesis **23**, 1229–34.
62. Kim, J.I., Y.J. Park, K.H. Kim, B.J. Song, M.S. Lee, C.N. Kim, and S.H. Chang. (2003). hOGG1 Ser326Cys polymorphism modifies the significance of the environmental risk factor for colon cancer. World J Gastroenterol **9**, 956–60.
63. Wang, Y., M.R. Spitz, Y. Zhu, Q. Dong, S. Shete, and X. Wu. (2003). From genotype to phenotype: correlating XRCC1 polymorphisms with mutagen sensitivity. DNA Rep **2**, 901–8.
64. Matullo, G., S. Guarrera, S. Carturan, M. Peluso, C. Malaveille, L. Davico, A. Piazza, and P. Vineis. (2001). DNA repair gene polymorphisms, bulky DNA adducts in white blood cells and bladder cancer in a case-control study. Int J Cancer 92(4), 562–7.
65. Hu, J.J., T.R. Smith, M.S. Miller, H.W. Mohrenweiser, A. Golden, and L.D. Case. (2001). Amino acid substitution variants of APE1 and XRCC1 genes associated with ionizing radiation sensitivity. Carcinogenesis **22**, 917–22.
66. Hu, J.J., T.R. Smith, M.S. Miller, K. Lohman, and L.D. Case. (2002). Genetic regulation of ionizing radiation sensitivity and breast cancer risk. Environ Mol Mutagen **39**, 208–15.
67. Lunn, R.M., D.A. Bell, J.L. Mohler, and J.A. Taylor. (1999). Prostate cancer risk and polymorphism in 17 hydroxylase (CYP17) and steroid reductase (SRD5A2). Carcinogenesis **20**, 1727–31.
68. Lunn, R.M., R.G. Langlois, L.L. Hsieh, C.L. Thompson, and D.A. Bell. (1999). XRCC1 polymorphisms: effects on aflatoxin B1-DNA adducts and glycophorin A variant frequency. Cancer Res **59**, 2557–61.
69. Matullo, G., D. Palli, M. Peluso, S. Guarrera, S. Carturan, E. Celentano, V. Krogh, A. Munnia, R. Tumino, S. Polidoro, A. Piazza, and P. Vineis. (2001). XRCC1, XRCC3, XPD gene polymorphisms, smoking and (32)P-DNA adducts in a sample of healthy subjects. Carcinogenesis **22**, 1437–45.
70. Fan, J., M. Otterlei, H.K. Wong, A.E. Tomkinson, and D.M. Wilson, 3rd. (2004). XRCC1 co-localizes and physically interacts with PCNA. Nucleic Acids Res **32**, 2193–201.
71. Tuimala, J., G. Szekely, S. Gundy, A. Hirvonen, and H. Norppa. (2002). Genetic polymorphisms of DNA repair and xenobiotic-metabolizing enzymes: role in mutagen sensitivity. Carcinogenesis **23**, 1003–8.
72. Xi, T., I.M. Jones, and H.W. Mohrenweiser. (2004). Many amino acid substitution variants identified in DNA repair genes during human population screenings are predicted to impact protein function. Genomics **83**, 970–9.
73. Hadi, M.Z., M.A. Coleman, K. Fidelis, H.W. Mohrenweiser, and D.M. Wilson, 3rd.

(2000). Functional characterization of Ape1 variants identified in the human population. Nucleic Acids Res **28**, 3871–9.

74. Dantzer, F., V. Schreiber, C. Niedergang, C. Trucco, E. Flatter, G. De La Rubia, J. Oliver, V. Rolli, J. Menissier-de Murcia, and G. de Murcia. (1999). Involvement of poly(ADP-ribose) polymerase in base excision repair. Biochimie **81**, 69–75.
75. Wieler, S., J.P. Gagne, H. Vaziri, G.G. Poirier, and S. Benchimol. (2003). Poly(ADP-ribose) polymerase-1 is a positive regulator of the p53-mediated G1 arrest response following ionizing radiation. J Biol Chem **278**, 18914–21.
76. Caldecott, K.W., C.K. McKeown, J.D. Tucker, S. Ljungquist, and L.H. Thompson. (1994). An interaction between the mammalian DNA repair protein XRCC1 and DNA ligase III. Mol Cell Biol. **14**, 68–76.
77. Sancar, G.B., W. Siede, and A.A. van Zeeland. (1996). Repair and processing of DNA damage: a summary of recent progress. Mutat Res **362**, 127–46.
78. Yu, M.W., S.Y. Yang, I.J. Pan, C.L. Lin, C.J. Liu, Y.F. Liaw, S.M. Lin, P.J. Chen, S.D. Lee, and C.J. Chen. (2003). Polymorphisms in XRCC1 and glutathione S-transferase genes and hepatitis B-related hepatocellular carcinoma. J Natl Cancer Inst **95**, 1485–8.
79. Spitz, M.R., X. Wu, Y. Wang, L.E. Wang, S. Shete, C.I. Amos, Z. Guo, L. Lei, H. Mohrenweiser, and Q. Wei. (2001). Modulation of nucleotide excision repair capacity by XPD polymorphisms in lung cancer patients. Cancer Res **61**, 1354–7.
80. Baccarelli, A., D. Calista, P. Minghetti, B. Marinelli, B. Albetti, T. Tseng, M. Hedayati, L. Grossman, G. Landi, J.P. Struewing, and M.T. Landi. (2004). XPD gene polymorphism and host characteristics in the association with cutaneous malignant melanoma risk. Br J Cancer **90**, 497–502.
81. Benhamou, S., and A. Sarasin. (2005). ERCC2 /XPD gene polymorphisms and lung cancer: a HuGE review. Am J Epidemiol **161,** 1–14.
82. Hou, S.M., S. Falt, S. Angelini, K. Yang, F. Nyberg, B. Lambert, and K. Hemminki. (2002). The XPD variant alleles are associated with increased aromatic DNA adduct level and lung cancer risk. Carcinogenesis **23**, 599–603.
83. Lunn, R.M., K.J. Helzlsouer, R. Parshad, D.M. Umbach, E.L. Harris, K.K. Sanford, and D.A. Bell. (2000). XPD polymorphisms: effects on DNA repair proficiency. Carcinogenesis **21**, 551–5.
84. Duell, E.J., J.K. Wiencke, T.J. Cheng, A. Varkonyi, Z.F. Zuo, T.D. Ashok, E.J. Mark, J.C. Wain, D.C. Christiani, and K.T. Kelsey. (2000). Polymorphisms in the DNA repair genes XRCC1 and ERCC2 and biomarkers of DNA damage in human blood mononuclear cells. Carcinogenesis **21**, 965–71.
85. Kiyohara, C., and K. Yoshimasu. (2007). Genetic polymorphisms in the nucleotide excision repair pathway and lung cancer risk: a meta-analysis. Int J Med Sci **4,** 59–71.
86. Huang, W.Y., S.I. Berndt, D. Kang, N. Chatterjee, S.J. Chanock, M. Yeager, R. Welch, R.S. Bresalier, J.L. Weissfeld, and R.B. Hayes. (2006). Nucleotide excision repair gene polymorphisms and risk of advanced colorectal adenoma: XPC polymorphisms modify smoking-related risk. Cancer Epidemiol Biomarkers Prev **15**, 306–11.
87. Mohrenweiser, H.W., D.M. Wilson, 3rd, and I.M. Jones. (2003). Challenges and complexities in estimating both the functional impact and the disease risk associated with the extensive genetic variation in human DNA repair genes. Mutat Res **526**, 93–125.
88. Araujo, F.D., A.J. Pierce, J.M. Stark, and M. Jasin. (2002). Variant XRCC3 implicated in cancer is functional in homology-directed repair of double-strand breaks. Oncogene **21,** 4176–80.
89. Nonoyama, S., and H.D. Ochs. (1996). Immune deficiency in SCID mice. Int Rev Immunol **13**, 289–300.
90. Nicolas, N., D. Moshous, M. Cavazzana-Calvo, D. Papadopoulo, R. de Chasseval, F. Le Deist, A. Fischer, and J.P. de Villarta. (1998). A human severe combined immunodeficiency (SCID) condition with increased sensitivity to ionizing radiations and impaired V(D)J rearrangements defines a new DNA recombination/repair deficiency. J Exp Med **188,** 627–34.
91. Dip, R. and H. Naegeli. (2005). More than just strand breaks: the recognition of structural DNA discontinuities by DNA-dependent protein kinase catalytic subunit. FASEB J **19,** 704–15.
92. Sipley, J.D., J.C. Menninger, K.O. Hartley, D.C. Ward, S.P. Jackson, and C.W. Anderson. (1995). Gene for the catalytic subunit of the human DNA-activated protein kinase maps to the site of the XRCC7 gene on chromosome 8. Proc Natl Acad Sci U S A **92,** 7515–9.
93. Zhang, Y., J. Zhou, and C.U. Lim. (2006). The role of NBS1 in DNA double strand

break repair, telomere stability, and cell cycle checkpoint control. Cell Res **16,** 45–54.

94. Medina, P.P., S.A. Ahrendt, M. Pollan, P. Fernandez, D. Sidransky, and M. Sanchez-Cespedes. (2003). Screening of homologous recombination gene polymorphisms in lung cancer patients reveals an association of the NBS1–185Gln variant and p53 gene mutations. Cancer Epidemiol Biomarkers Prev **12**, 699–704.
95. Margison, G.P., A.C. Povey, B. Kaina, and M.F. Santibanez Koref. (2003). Variability and regulation of O6-alkylguanine-DNA alkyltransferase. Carcinogenesis **24**, 625–35.
96. Inoue, R., M. Abe, Y. Nakabeppu, M. Sekiguchi, T. Mori, and T. Suzuki. (2000). Characterization of human polymorphic DNA repair methyltransferase. Pharmacogenetics **10**, 59–66.
97. Margison, G.P., J. Heighway, S. Pearson, G. McGown, M.R. Thorncroft, A.J. Watson, K.L. Harrison, S.J. Lewis, K. Rohde, P.V. Barber, P. O'Donnell, A.C. Povey, and M.F. Santibanez-Koref. 92005). Quantitative trait locus analysis reveals two intragenic sites that influence O6-alkylguanine-DNA alkyltransferase activity in peripheral blood mononuclear cells. Carcinogenesis **26**, 1473–80.
98. Matsuoka, S., G. Rotman, A. Ogawa, Y. Shiloh, K. Tamai, and S.J. Elledge. (2000). Ataxia telangiectasia-mutated phosphorylates Chk2 in vivo and in vitro. Proc Natl Acad Sci USA **97**, 10389–94.
99. Kim, J.H., H. Kim, K.Y. Lee, K.H. Choe, J.S. Ryu, H.I. Yoon, S.W. Sung, K.Y. Yoo, and Y.C. Hong. (2006). Genetic polymorphisms of ataxia telangiectasia mutated affect lung cancer risk. Hum Mol Genet **15**, 1181–6.
100. Koren, M., G. Kimmel, E. Ben-Asher, I. Gal, M.Z. Papa, J.S. Beckmann, D. Lancet, R. Shamir, and E. Friedman. (2006). ATM haplotypes and breast cancer risk in Jewish high-risk women. Br J Cancer **94**, 1537–43.
101. Lee, K.M., J.Y. Choi, S.K. Park, H.W. Chung, B. Ahn, K.Y. Yoo, W. Han, D.Y. Noh, S.H. Ahn, H. Kim, Q. Wei, and D. Kang. (2005). Genetic polymorphisms of ataxia telangiectasia mutated and breast cancer risk. Cancer Epidemiol Biomarkers Prev **14**, 821–5.
102. Audebert, M., J.P. Radicella, and M. Dizdaroglu. (2000). Effect of single mutations in the OGG1 gene found in human tumors on the substrate specificity of the Ogg1 protein. Nucleic Acids Res **28**, 2672–8.
103. Chen, C., N.S. Weiss, F.Z. Stanczyk, S.K. Lewis, D. DiTommaso, R. Etzioni, M.J. Barnett, and G.E. Goodman. (2003). Endogenous sex hormones and prostate cancer risk: a case-control study nested within the Carotene and Retinol Efficacy Trial. Cancer Epidemiol Biomarkers Prev **12**, 1410–6.
104. Peng, T., H.M. Shen, Z.M. Liu, L.N. Yan, M.H. Peng, L.Q. Li, R.X. Liang, Z.L. Wei, B. Halliwell, and C.N. Ong. (2003). Oxidative DNA damage in peripheral leukocytes and its association with expression and polymorphisms of hOGG1: a study of adolescents in a high risk region for hepatocellular carcinoma in China. World J Gastroenterol **9**, 2186–93.
105. Khan, S.G., E.J. Metter, R.E. Tarone, V.A. Bohr, L. Grossman, M. Hedayati, S.J. Bale, S. Emmert, and K.H. Kraemer. (2000). A new xeroderma pigmentosum group C poly(AT) insertion/deletion polymorphism. Carcinogenesis **21**, 1821–5.
106. Wang, C.Y., R.F. Jones, M. Debiec-Rychter, G. Soos, and G.P. Haas. (2002). Correlation of the genotypes for N-acetyltransferases 1 and 2 with double bladder and prostate cancers in a case-comparison study. Anticancer Res **22**, 3529–35.

Chapter 19

Linking the Kaposi's Sarcoma-Associated Herpesvirus (KSHV/HHV-8) to Human Malignancies

Inna Kalt, Shiri-Rivka Masa, and Ronit Sarid

Summary

In 1994, the Kaposi's sarcoma-associated herpesvirus (KSHV/HHV-8) was identified as the etiologic agent of Kaposi's sarcoma (KS). KSHV has since been associated with two additional malignancies: primary effusion lymphoma and multicentric Castleman's disease. In this chapter, we describe the current understanding of the pathogenesis, transmission, and prevalence of KSHV, and its association mainly with KS. We describe evidence demonstrating that KSHV is a causative agent for KS, and we present other factors that possibly contribute to the incidence of KS. We compare worldwide data on the prevalence of KS and of KSHV infection. Specific viral genes that may induce KS tumors or enable their growth also are described. Finally, we discuss the implications of the transmission modes and epidemiology of this virus on recommendations for KSHV screening of tissues and blood products before transplantation or transfusion.

Key Words: Kaposi's sarcoma-associated herpesvirus (KSHV), human herpesvirus 8 (HHV-8), Kaposi's sarcoma (KS), primary effusion lymphoma (PEL), multicentric Castleman's disease (MCD).

1. Introduction: Discovery of Kaposi's Sarcoma-Associated Herpesvirus

Kaposi's sarcoma (KS) is a complex multifocal neoplasm, characterized by proliferation of spindle-like cells, substantial angiogenesis, and inflammation. KS was originally described in 1872 by the Hungarian dermatologist Moritz Kaposi as an idiopathic, multiple-pigmented sarcoma of the skin, with a fatal outcome *(1)*. This type of KS, which later became known as classic KS, occurs mainly in elderly people, particularly men of Mediterranean, Eastern European, or Jewish heritage. However, in contrast to its first description, classic KS typically follows an indolent course, and it tends to evolve slowly, sometimes over decades, usually starting on the feet and primarily involving the skin

Mukesh Verma (ed.), *Methods in Molecular Biology, Cancer Epidemiology, vol. 471*
© 2009 Humana Press, a part of Springer Science + Business Media, Totowa, NJ
Book doi: 10.1007/978-1-59745-416-2

(2–4). A high occurrence of African or endemic KS was reported during the early 1950s in equatorial African countries. Endemic KS takes a form similar to classic KS in immunocompetent individuals, but it affects children as well, in a progressive lymphadenopathic disease that often has a poor prognosis *(5–7)*. During the 1960s, KS emerged among immunosuppressed patients after solid-organ transplantation. This iatrogenic form tends to be clinically aggressive, underscoring the importance of the immune system in modulating the progression of the disease. Accordingly, discontinuation of immunosuppressive therapy is often associated with clinical remission *(8,9)*. A tremendous increase in the incidence of KS was noticed during the early 1980s, concomitantly with the emergence of the acquired immunodeficiency syndrome (AIDS). KS was noticed as one of the first clinical manifestations of human immunodeficiency virus (HIV) infection, and it became a main characteristic of AIDS, leading to the characterization of KS as an AIDS-associated malignancy *(10)*. The clinical course of AIDS-KS is variable, ranging from localized and indolent to widespread, involving lymphatic and oral tissues. AIDS-KS tends to spread to visceral organs, and this disseminated disease has a fatal outcome. Since the introduction of highly-active antiretroviral therapy (HAART), the incidence of KS among AIDS patients has declined considerably *(11–13)*.

At present, KS remains the most frequent neoplasm in AIDS patients, whereas the increasing rates of both noncompliance and treatment failure in HAART patients suggest that KS will continue to represent a significant public health concern. In African countries, the high prevalence of HIV infection has resulted in the continuous rise in the incidence rates of KS, which has become the most common cancer in children and adult men in certain countries. Furthermore, there is an increasing concern regarding the development of KS in solid-organ transplant patients, who receive immunosuppressive therapy.

The etiology of KS, in all its epidemiologic and clinical settings, has been questioned for over 100 years. The geographic variability in the incidence of KS together with the relatively high occurrence of AIDS-KS, particularly among homosexual men, suggested the existence of a unique "KS infectious agent" *(14,15)*. Significant efforts were made to identify the agent responsible for KS, and several candidates including cytomegalovirus, mycoplasma, papillomavirus, and BK virus were suggested to play an etiologic role in the disease; however, none was systematically identified in KS patients. The particularly aggressive form of KS in AIDS patients encouraged scientists to investigate the role of the HIV-1 Tat protein in KS growth *(16,17)*. However, the observation that KS is mainly observed in homosexual AIDS patients, and less in individuals who acquired HIV-1 through blood transfusion or

intravenous drug use, did not support the angiogenic features of Tat protein as the major cause of KS.

In 1994, by using a subtractive hybridization technique (representational difference analysis), Chang and colleagues *(18)* amplified two novel DNA sequences from an AIDS-associated KS lesion. These sequences were absent in DNA from unaffected tissue from the same patient, and displayed homology to herpesviral capsid and tegument genes. Based on the primary association of the newly discovered virus to KS, the virus containing these genes was named Kaposi's sarcoma-associated herpesvirus (KSHV) or human herpesvirus 8 (HHV-8) according to formal herpesvirus nomenclature. The two isolated fragments enabled the isolation of contiguous fragments, which ultimately led to genomic cloning and complete sequencing of the 165-kbp double-stranded DNA viral genome *(19)*. Thus, the discovery of KSHV emerged from epidemiologic data that suggested a sexually transmitted contagious agent, which led its identification using molecular techniques. The importance of interdisciplinary partnerships, leading to this type of scientific progress, is exemplified by the fact that the husband and wife team who discovered KSHV, Yuan Chang and Patrick Moore, are a pathologist and a public health epidemiologist.

Soon after its discovery, KSHV was established as the primary causative factor in all types of KS, in immunocompromised as well as immunocompetent individuals *(20)*. KSHV also has been linked to the pathogenesis of certain lymphoproliferative disorders, principally primary effusion lymphoma (PEL, also termed body cavity-based lymphoma) *(21)* and some cases of multicentric Castleman's disease (MCD) *(22)*. These malignancies are observed most frequently among HIV patients, but they also have been documented in transplant patients, and, to a lesser degree, in apparently immunocompetent individuals. Rare cases exhibiting a concomitant occurrence of two or three of the KSHV-related diseases have been reported in the literature, supporting a common causative agent *(23–25)*. The involvement of KSHV with other disease conditions has been widely investigated *(3)*. To date, the association of KSHV with a broad spectrum of diseases has been rejected, although the occasional involvement in other pathologies in unique subsets of clinical settings remains possible.

Extensive epidemiological and virological studies have resulted in further elucidation of the serological profiles, modes of viral transmission, molecular organization, and potential tumorigenic pathways induced by KSHV infection. Obviously, this can provide unique insight into the mechanisms of cancer in general, and of virus-associated malignancies in particular.

2. Natural KSHV Infection

As a member of the gammaherpesvirus subfamily, KSHV shares biological characteristics with other herpesviruses, including the capacity for life-long persistence after primary infection. The characteristics of primary or acute infection with KSHV have been incompletely defined. A febrile illness that was sometimes associated with maculopapular skin rash and active KSHV viremia was described in immunocompetent children *(26–28)*. Primary KSHV infection in immunocompetent adults is usually associated with mild diarrhea, fatigue, localized rash, and lymphadenopathy *(29)*. Fever and cutaneous rash, coinciding with KSHV viremia, have been reported as part of primary infection in immunosuppressed patients after transplantation *(30)*. A severe, sometimes fatal, outcome (e.g., bone marrow or multiorgan failure) of primary infection was described in rare cases of immunocompromised patients *(31,32)*.

During lifetime infection, KSHV displays two distinct cycles: lytic (productive) and latent. Extensive viral DNA replication and a well-controlled sequence of viral gene expression characterize the lytic cycle, which may culminate in the production of infectious progeny virus and host cell death *(33–35)*. This cycle is thought to take place during primary infection, and it is crucial for virus spread between cells and hosts. It is also likely to play an important role in the tumorigenesis induced by KSHV *(36)*. In contrast to the lytic cycle, during latency, the viral genome persists as a circular DNA (mini-chromosome/episome); few viral genes are expressed, and no viral particles are produced. The cellular site of KSHV latency and the mechanisms by which the latent virus escapes elimination by the host immune system are not fully understood. Certain physiologic conditions may periodically reactivate the hidden latent virus in most asymptomatic carriers, increasing the risk for disease onset. Only some of the variables involved in this switch are known, and the interplay between these variables is even less well understood *(37–47)*. The role of host immunodeficiency and the impact of the HIV-encoded TAT protein on KSHV activation and pathogenicity are well established *(48)*. Accordingly, the use of HAART has led to a decreased incidence of KS in AIDS patients and to the remission of clinical KS in treated patients *(11–13)*. Similarly, KSHV reactivation and KS onset have been widely documented upon immunosuppressive treatment in posttransplant patients *(9,49)*. In the context of immunodeficiency, the level of KSHV viremia is strongly associated with the onset of KSHV-related disease *(50–54)*. However, additional important factors and host signal transduction pathways that apparently operate in the

transition between the asymptomatic viral carrier state, and the development of KSHV-associated diseases remain to be elucidated.

The development of malignancies in KSHV-infected individuals is fairly rare, pointing to the involvement of additional cofactors. The most frequent KSHV-associated malignancy is KS *(3,4)*. KS may develop in various organs but, most usually, lesions are seen on the skin. Pure cutaneous involvement usually has a benign course, whereas visceral dissemination, which is more common in AIDS patients, can be aggressive. The KS lesions contain both spindle-shaped endothelial cells that are latently infected with KSHV, and cells that support the production of progeny virus *(55)*. In addition to spindle cells, all KS lesions contain variable inflammatory lymphocytic infiltrates including monocytes, T cells, B cells, histiocytes, and plasma cells. Recent evidence suggests that KSHV fails to provide a proliferative advantage to the latently infected cells; therefore, these latently infected cells are rapidly lost. Consequently, the frequent productive transcription program observed in spindle cells within KS lesions is necessary to maintain virus infection *(56–59)*. Elevated levels of viral DNA in peripheral blood mononuclear cells (PBMCs) have been reported among KS patients and in individuals subsequently progressing to KS, compared with healthy subjects. Yet, the relationship between the viral load in PBMCs and the extent of the disease is subject to debate *(50–54,54,60)*.

The etiologic association between PEL and KSHV has mainly relied on cases observed in HIV-infected patients, whereas only a small number of cases have been described in HIV-negative elderly patients. Of patients with PEL, 25% also have KS *(61)*. PEL cells are of B-cell lineage, possessing immunoglobulin rearrangements, and commonly lacking T-cell and most B-cell surface markers. Almost all cases of PEL harbor KSHV, principally in a latent phase, with 40–150 episomal copies of the virus genome per cell. The scarcity of PEL suggests that KSHV infection represents only one of the several events participating in the cascade that leads to the development of PEL. Because most PEL tumors are co-infected with Epstein–Barr virus (EBV) as well, EBV infection was included among the cofactors *(62)*.

Like PEL, MCD occurs at increased frequency in AIDS patients, and on this background is almost always associated with KSHV infection. MCD also may occur in HIV-negative individuals, yet only 40 to 50% of these cases are KSHV positive. In MCD, the cells harboring KSHV occur in the mantle zone of B-cell follicles, eventually forming microscopic foci of clonal plasmablastic lymphoma. MCD is associated with abundant lytic infection, suggesting the involvement of viral lytic proteins in the pathogenesis of this disease *(22,63)*.

3. KSHV Transmission

More than 95% of KS skin lesions test positive for KSHV DNA. KSHV DNA has been found in the PBMCs of the majority of individuals with AIDS-KS and in 8–10% of healthy subjects positive for KSHV *(64)*. Healthy KSHV-infected individuals and KS patients periodically shed infectious viruses from the oral epithelial cells *(65)* and from the oropharynx to the saliva *(66–68)*. Semen and prostatic fluid from KSHV-infected individuals contain low amounts of infected cells and viral DNA *(69,70)*. Accordingly, transmission of KSHV has been observed by means of various forms of interactions, including mother to child, sexual contact, household, blood transfusion, needle sharing, and organ transplantation. Because both human epithelial cells and keratinocytes support KSHV infection and replication *(71–73)*, mucosal tissues are likely to be the most important site of infection. Nonetheless, the predominant transmission modalities of KSHV seem to differ between geographic areas and populations.

Among homosexual men, KSHV infection is associated with a high lifetime number of sexual partners, other sexually transmitted diseases, and sexual contact with other men from communities where AIDS-KS is common. Thus, sexual transmission, which can probably occur through oral–genital, anal–genital, and oral–anal intercourse seems to play an important role among homosexual men *(74–78)*. However, because KSHV is not substantially shed into the semen or rectal tissue, the reason for the predominance of sexual transmission in this particular population and the specific route of virus transmission remain unresolved.

The risk factors for KSHV acquisition in heterosexual individuals are not yet fully understood. Several studies in African countries demonstrated an association between KSHV-seropositive status, and multiple sex partners and sexually transmitted diseases. However, in these regions, KSHV is frequently acquired in childhood through horizontal transmission, and the prevalence of KSHV steadily increases with age, in the absence of sexual exposure *(27,79–82)*. This supports an additional, nonsexual horizontal transmission route. The high prevalence of infection before puberty in these areas and the high detection rates of KSHV in saliva provide further indirect evidence for nonsexual transmission. Mother-to-child transmission, during pregnancy or at delivery, is also likely to occur in endemic areas, but the low KSHV seroprevalence in children younger than 5 years argues against this route as a major mode of virus dissemination. Furthermore, no evidence of KSHV DNA was found in amniotic fluid or in cord blood of Italian pregnant women in whom KSHV DNA was detected *(83)*, providing direct evidence that vertical transmission

of KSHV is uncommon. In North America and Western Europe, infection is rare in childhood and the acquisition of KSHV infection occurs mainly after puberty. However, sexual activity has not been specifically associated with an increased risk of KSHV infection in this heterosexual population. Likewise, in the developing world, most children become infected with EBV within the first decade of life, whereas in the more developed Western world, up to half of children remain seronegative at the end of their first decade and become infected in adolescence or young adulthood. Together, horizontal transmission, predominantly through the saliva, seems to be the foremost route for KSHV spread.

Within family members of classic and endemic KS patients, the prevalence of KSHV is considerably higher than in the corresponding general population *(84,85)*. Similarly, significant correlations between the KSHV serostatus of mothers and offspring, between that of siblings, and between that of spouses were reported in families from the Mediterranean and African countries not affected with KS *(86,87)*. This further supports salivary shedding of KSHV as a possible mode of intrafamilial transmission, where exposure to saliva may occur during play, through shared eating implements, and through kissing. The high correlation between seropositivity of the mother and that of the offspring further supports this notion, and it could suggest that regular maternal activities could result in transmission to other family members.

The extent to which KSHV is transmitted by blood transfusion remains controversial. Blood or blood products can potentially transmit KSHV, because viral DNA is found in PBMCs. However, this risk is generally estimated to be low, and it is dependent upon KSHV prevalence in specific populations. Even among HIV-positive patients, the prevalence of KSHV in hemophilia patients, intravenous drug users, and transfusion recipients is low. It was suggested that the transmissibility of the virus via the blood may be limited by its cell-associated nature and its low load in asymptomatic seropositive individuals. Nevertheless, two recent comprehensive studies provided strong evidence for transmission of KSHV by blood transfusion *(88,89)*. In countries that are highly endemic for KSHV, transmission by blood is not thought to be a major modality for KSHV spread. In low-prevalence areas, this risk is similarly low, and transmission by blood is thought to occur rarely, although it is feasible.

KSHV can be transmitted during organ transplantation through an infected allograft. Reactivation of a pre-existing virus in the transplant recipient may account for KS development after organ transplantation; however, the risk for morbidity and mortality increases among patients who acquire primary infection during organ transplantation. The proportion of KS cases occurring by primary infection versus virus reactivation varies according to the prevalence of KSHV infection *(9)*. Interestingly, specific

donor-derived genetic markers were detected in cells inhabiting KS lesions that were isolated from renal transplant patients *(90)*. This suggests that the infected organ might not only transmit the virus but also may seed progenitor cells that persist in the recipients and can undergo transformation and progression.

4. Global Prevalence of KSHV Infection and KS

The incidence of KS varies among countries and populations, whereas incidence comparison is impeded by the difficulty in distinguishing non-HIV from HIV-associated KS. Generally, the incidence rates of KS range, in men, from less than 0.5/100,000 in Asia and Northern Europe, to 2.7–3 in Israel and certain parts of Italy and 50.8/100,000 in Harare, Zimbabwe (**Fig. 19.1A**; (*91*)). Rates of KS disease increase up to 500–1000-fold in solid organ transplant patients and up to 20,000-fold in male homosexual AIDS patients. Thus, in the United States, the incidence rate of KS is fairly low, although it rises among HIV-positive individuals. The large increase in KS in southern Africa, where it is now the most frequently encountered cancer, reflects the high prevalence of both HIV and KSHV infection.

Antibody response persists after primary infection, and it can be used to establish the prevalence of infection and to evaluate risk factors for transmission. Hence, the development of serologic assays to detect antibodies against KSHV has paved the way for large-scale epidemiologic studies *(92–95)*. At present, no single assay based on recombinant viral antigens has sufficient sensitivity and specificity to identify infected individuals in routine clinical settings. A higher predictive value is achievable by the use of multiple serologic assays using both latent and lytic viral antigens *(96)*. Despite these limitations, the relative prevalence rates of KSHV reported for different populations are fairly comparable and demonstrate distinct trends (**Fig. 19.1B**).

Unlike most other human herpesviruses, KSHV is not globally ubiquitous. Almost all patients with KS have antibodies to KSHV, independent of their ethnicity and their country of residence. Seropositivity rates for KSHV show remarkable racial and geographic variations (**Fig. 19.1B**). Accordingly, the global prevalence of KSHV infection can be categorized into four major patterns of prevalence: very high (>40%) in sub-Saharan Africa, high (between 20 and 40%) in southern Italy and certain parts of Africa, moderate (between 5 and 20%) in Italy and other Mediterranean countries, and low (up to 5%) in Northern and Western Europe and in the United States *(3,97–99)*. No differences in the prevalence KSHV-positive individuals were described between

Fig. 19.1. Epidemiologic patterns of KSHV infection and KS. Countries were classified on the basis of available data on the incidence rates of KS in males/100,000 individuals/year **(A)** and on the prevalence of KSHV infection **(B)**. In most geographical areas, the overall incidence rates of KS do not distinguish between the different epidemiologic and clinical settings of KS. This is exemplified in the United States, where the KSHV seroprevalence and KS incidence are generally low, whereas in cities that have large homosexual communities, such as San Francisco and Los Angeles, the incidence rates of KS are high. The KSHV seroprevalence map combines data from numerous studies on HIV-negative populations.

men and women in different countries. However, KSHV infection rates differ widely between HIV-positive and -negative populations, and may differ within small areas. Among HIV exposure subgroups in North America and Europe, the prevalence of KSHV has distinct trends. Within a defined geographical region, this prevalence is highest among HIV-infected homosexual men, lower among HIV-infected injecting drug users or HIV heterosexual individuals, and the lowest KSHV prevalence is among HIV-infected children. This trend mirrors the pattern of KS incidence in HIV-infected individuals. Interestingly, segregation analysis of KSHV seroprevalence rates in French Guyana suggested the existence of a major recessive gene predisposing individuals to KSHV infection with gene-by-age interaction *(100)*.

Generally, a strong direct correlation between the seroprevalence of KSHV and the incidence rates of KS has been confirmed by several comparisons of data on the worldwide level. A few notable exceptions to this rule have been reported. In Egypt *(28)*, Gambia *(101)*, certain areas of China *(102)*, and among Brazilian-Amerindians *(103)* and Ethiopian immigrant population in Israel *(104)*, the seroprevalence of KSHV is high, whereas the incidence of KS is unexpectedly low. Conversely, several cases of familial KS have been reported *(105,106)*. These unique cases suggest that unidentified cofactors may protect or expose certain populations from the clinical consequences of viral infection.

5. Genomic Variants of KSHV

KSHV is an ancient virus thought to have coexisted alongside Homo sapiens since their origin *(107)*. Phylogenetic analyses revealed several distinct subtypes of KSHV, which are thought to have diverged at least 100,000 years ago, with additional splits occurring later. Based on variation in the left-most open reading frame (ORF) K1 (VIP) gene, the first subtypes to be identified were A, B, C, and D, followed by subtype E. Additional variations were identified in the right end of the genome, in ORF K15 (TMP). K15 genes fall into two alternative allelic subtypes, referred to as prototype (P) and minor (M), which have diverged by 70% at the amino acid level. The variability in ORF K15 is not linked to that of ORF K1. In addition, small variations have been identified in 10 internal genomic loci, which together with VIP and TMP variations, establish a series of 12 major genotype patterns.

Hayward and colleagues *(107)*, who conducted most of the studies on KSHV genomic variants, concluded that the principal K1 subtypes arose during the migration of modern humans out of east Africa first into sub-Saharan Africa (variant B), then into south Asia, Australia, and the Pacific Rim (variants D and E), and finally into the Middle East, Europe and north Asia (variants A and C),

with very little subsequent mixing. Accordingly, subtypes A and C predominate in Europe, whereas subtype B predominates in Africa. Subtype D is rare and has been found in individuals of Polynesian and Australian aboriginal descent. Subtype E has been reported to be hyperendemic in Brazilian Amerindians. Interestingly, 20% of all human KSHV genomes investigated so far contain the K15 M allele, but this allele can be found associated with all K1 gene variants.

Viral genome analysis can be applied to track the origin and modes of virus infection, and it has been used to establish inter- and intrafamilial transmission of KSHV as well as to follow genotypic differences in virus isolated from different body compartments of a single patient.

6. Causal Association between KSHV and KS and Risk Factors for KS

KSHV is now considered a prerequisite etiologic agent for all variants of KS, including the classic, endemic, iatrogenic, and epidemic KS types *(108)*. This widely accepted notion is based on consistent and reproducible evidence that was obtained by numerous investigators. Specifically, 1) KSHV DNA is commonly detected in KS lesions but not in the adjacent tissue of patients with all forms of the disease; 2) KSHV infection precedes the onset of KS. Furthermore, detection of antibodies directed against KSHV antigens in the blood of HIV-infected subjects is strongly predictive of their future development of KS; 3) a general correlation between KSHV prevalence and population risk for developing KS is well established; 4) KSHV is present in critical cellular components of KS, namely, the endothelial and spindle cells; and 5) KSHV encodes several potentially oncogenic proteins and it is closely related to other tumorigenic viruses. Yet, KSHV infection is not sufficient to induce KS, and other host-related factors are probably involved in the pathogenesis of the disease, because only a small minority of KSHV-infected individuals will actually develop KS or other virus-associated diseases. Classic KS was estimated to develop annually in only 0.03% of KSHV-infected men, and in 0.01–0.02% of the KSHV-infected women above age 50 *(109)*. The incidence of KS among posttransplant immunosuppressed patients is 500–1,000 times greater than that of the general population *(110)*, whereas untreated HIV patients possess the highest risk for KS. The risk of KS among HIV patients who acquire KSHV infection is estimated to be up to 20,000-fold higher compared with healthy KSHV-infected individuals. This not only illustrates the prominence of immunological performance in the development of KS but also reflects specific interactions between HIV and KSHV proteins that yield a synergistic effect. Of note, the presence of HIV infection before KSHV acquisition predicts a more rapid progression to KS *(111)*. This may be due to the enhancement of KSHV pathogenesis in the presence of HIV, or it may reflect an impaired immune response to KSHV in HIV-infected individuals.

Several genetic, immune, and environmental cofactors have been suggested to affect the risk of KS development after KSHV acquisition. Age is an established cofactor for the development of KS among KSHV-positive individuals. Except in endemic areas for KSHV, KS in immunocompetent individuals usually develops after 50 years of age. Whether the early age of KSHV acquisition in the Africa is a contributing factor to KS development in children in this area is unknown. In addition, the reasons for the tendency of KS to selectively affect men are unknown. Genetic determinants, possibly those that reflect ethnicity , such as the human leukocyte antigen type may be influential *(112–114)*. Selected alleles of the interleukin-6 promoter *(115)* and of the low-affinity Fcγ receptors IIIA for IgG *(116)* are associated with an increased lifetime risk of the development of KS in men infected with HIV. Among KSHV-seropositive Italians, classic KS risk is associated with diplotypes of IL8RB and the promoter region of IL13 *(117)*. Underlying immune activation, such as elevated level of neopterin or β-2 microglobulin, and immunosuppression reflected in lower CD4 cell counts and low lymphocyte count also were found to be associated with the disease *(60)*.

Environmental factors probably play an important role, and have always drawn attention *(118)*. For example, KS is associated with exposure to volcanic soils, and it seems to be more common in highland areas located at altitudes >2,000 feet, although high incidence rates have been found at a broad range of altitudes and temperatures *(119)*. Lifestyle factors have been demonstrated to influence KS occurrence. Cigarette smoking is inversely associated with classic KS risk, whereas HIV-infected homosexual men in the United States who smoke also were found to have a reduced risk of AIDS-associated KS *(120,121)*. However, in other studies, tobacco use was not associated with classic, AIDS-associated or endemic KS *(122)*. Crack cocaine also has been inversely associated with KS *(120)*. Infrequent bathing was demonstrated to be associated with an increased risk of classic KS *(121)*. A history of several diseases was associated with an elevated risk of KS. These include malaria, asthma, and allergies *(121)*.

At present, it seems that the virulence of specific strains of KSHV is not likely to play an important role in their ability to induce KS. However, the effect of viral factors remains to be further explored.

7. Molecular Mechanisms of KSHV Pathogenesis

KSHV shares close homology with several primate and nonprimate mammalian viruses. Among human herpesviruses, it is most closely related to the EBV, known as a ubiquitous lymphotrophic

virus, which is associated causally with several human cancers, including nasopharyngeal carcinoma and certain types of lymphomas. KSHV encodes 12 microRNAs and >85 ORFs *(19,123–126)*, including several homologs of critical cellular proteins, likely pirated from the host. Unlike their cellular counterparts, the viral proteins escape cellular regulatory pathways and hence have altered activity *(127–129)*. The tumorigenic potential of certain KSHV-encoded genes has been revealed using in vitro assays, whereby their individual expression in rodent cells induces transformation. For some virally encoded genes, transformation in these assays necessitates coexpression of another cellular or viral oncogene. Detailed studies of the molecular mechanisms of transformation has indicated that certain viral genes act as proliferative, antiapoptotic or survival factors; thus, these genes probably contribute to KSHV-induced tumorigenesis by extending the lifespan of virus-infected cells. Furthermore, KSHV infection causes alterations in the cellular gene expression profile, both at the level of transcription and at the level of mRNA stability. In addition to these capacities, KSHV has evolved the ability to evade recognition and attack by host's innate and adaptive immune response *(130)*.

Justified by the notion that the productive virus infection culminates in cell death, initial efforts to identify KSHV genes responsible for its tumorigenic potential focused primarily on viral genes that are expressed during latent infection. These include the latency-associated nuclear antigen (LANA)-1, vCyclin, vFLIP, Kaposin, and LANA-2. The latent genes, except for LANA-2, which is exclusively expressed in PEL and MCD cells, are expressed in all KSHV-infected cells; therefore, they are predicted to play a role in KSHV oncogenesis. LANA-1 is a multifunctional protein, which tethers the viral episome to host chromatin during mitosis to ensure segregation of viral genomes to daughter cells. Similar to other oncogenes of DNA tumor viruses, LANA-1 binds to and interferes with the functions of the tumor suppressors p53 and retinoblastoma protein (pRB), and it transforms primary rat embryo cells when coexpressed with H-ras. vCyclin complexes with cellular cdk6 to phosphorylate a wide range of targets, including the pRB, p27, and histone H1, mimicking the effects of mitotic cyclins. vFLIP blocks apoptotic pathways and protects cells from Fas/APO1-mediated apoptosis by inhibiting activation of particular caspases and by activating nuclear factor-κB, which is essential for growth and survival of cells. Degradation of particular short-lived cytokine transcripts is blocked through activation of the p38-MK2 pathway by Kaposin B. This may lead to a cellular environment more compatible with transformation. LANA-2 promotes tumorigenesis through the inhibition of p53 and PKR-dependent apoptosis, and of the interferon-dependent and FOXO3a transcriptional activities.

Emerging evidence, however, indicates that the expression of latent proteins may not be sufficient to initiate and sustain KS, suggesting a requirement for viral lytic genes in the initiation of KSHV-related malignancies. Among the lytic genes, the KSHV G protein-coupled receptor (vGPCR) is the leading candidate gene responsible for the initiation of KS. vGPCR can activate various key intracellular molecules, including mitogen-activate protein kinase, p38, c-Jun NH_2-terminal kinase, and Akt, which in turn may control the expression and activity of numerous growth-promoting proteins. Because vGPCR induces the secretion of angiogenic growth factors, it was suggested to play a role both in direct autocrine and indirect paracrine maintenance of the transformed state *(59,131)*.

Thus, it seems that KS represents a consequence of a multistep process involving both viral and cellular factors that initiate reactive and inflammatory processes. Subsequently, some forms of selective pressure or genetic alteration, which are induced in latently infected cells, could give rise to the monoclonal form of KS. This view is supported by studies showing that KS lesions display all patterns of clonality (mono-, oligo-, and polyclonal), suggesting that KS begins as a polyclonal hyperplasia with potential for subsequent evolution to a monoclonal tumor.

8. Concluding Remarks

Tumor viruses cause at least 15% of human cancers. Globally, KS is the fourth most common cancer caused by infection, after gastric cancer (*Helicobacter pylori*), cervical cancer (human papillomavirus), and liver cancer (hepatitis viruses) *(132)*. Among immunosuppressed patients and patients who have undergone organ transplantation, KS remains a major cause of cancer-related deaths. The recognition that a virus may play an etiological role in a specific type of cancer can profoundly affect the ways in which the tumor is diagnosed, treated, and prevented. Furthermore, oncogenic viruses provide clues to how nonvirally induced cancers might develop.

Screening organ donors and recipients for KSHV infection is still controversial, and currently, KSHV is not included in routine virological assessment of donor organs or of transplant recipients. In light of seroepidemiologic studies, systematic screening for KSHV, not to exclude the graft, but to verify its KSHV status and to enable adequate monitoring of recipients is warranted. This may enable prophylaxis with antiviral agents or immunoglobulins to prevent virus reactivation in the recipient or from the graft. Consequently, the occurrence of severe KSHV-related diseases,

could be prevented, earlier diagnosis of cases that do develop KS could be accomplished, and non-KS morbidity associated with primary KSHV infection would decline. Furthermore, given the large numbers of blood transfusions that are given each year worldwide, careful consideration of the cost–benefit of the screening of blood products for KSHV also should take place. Because there is no clinically approved KSHV screening test, the development of a test is critical for controlling KSHV transmission and mitigating the consequences of infection.

Acknowledgments

We regret the omission of many research papers due to space constraints. We thank Ella Gindy for graphical assistance. Work in our laboratory is supported by the Israeli Ministry of Justice, the Israeli Ministry of Health, the Israel Association for Cancer Research, and by the Israel Science Foundation.

References

1. Kaposi, M. (1872) Idiopathisches Multiples Pigmentsarkom Der Haut. **3**, 265–273.
2. Iscovich, J., Boffetta, P., Franceschi, S., Azizi, E., and Sarid, R. (2000) Classic Kaposi sarcoma: epidemiology and risk factors. **88**, 500–517.
3. Cohen, A., Wolf, D. G., Guttman-Yassky, E., and Sarid, R. (2005) Kaposi's sarcoma-associated herpesvirus: clinical, diagnostic, and epidemiological aspects. **42**, 101–153.
4. Dourmishev, L. A., Dourmishev, A. L., Palmeri, D., Schwartz, R. A., and Lukac, D. M. (2003) Molecular genetics of Kaposi's sarcoma-associated herpesvirus (human herpesvirus-8) epidemiology and pathogenesis. **67**, 175–212.
5. Cook-Mozaffari, P., Newton, R., Beral, V., and Burkitt, D. P. (1998) The geographical distribution of Kaposi's sarcoma and of lymphomas in Africa before the aids epidemic. **78**, 1521–1528.
6. Sitas, F. and Newton, R. (2001) Kaposi's sarcoma in South Africa. 1–4.
7. Sissolak, G. and Mayaud, P. (2005) AIDS-related Kaposi's sarcoma: epidemiological, diagnostic, treatment and control aspects in sub-Saharan Africa. **10**, 981–992.
8. Zmonarski, S. C., Boratynska, M., Puziewicz-Zmonarska, A., Kazimierczak, K., and Klinger, M. (2005) Kaposi's sarcoma in renal transplant recipients. **10**, 59–65.
9. Marcelin, A. G., Calvez, V., and Dussaix, E. (2007) KSHV after an organ transplant: should we screen? **312**, 245–62.
10. Goedert, J. J. (2000) The epidemiology of acquired immunodeficiency syndrome malignancies. **27**, 390–401.
11. Cheung, T. W. (2004) AIDS-related cancer in the era of highly active antiretroviral therapy (HAART): a model of the interplay of the immune system, virus, and cancer. "On the offensive--the Trojan horse is being destroyed"--part A: Kaposi's sarcoma. **22**, 774–786.
12. Bower, M., Palmieri, C., and Dhillon, T. (2006) AIDS-related malignancies: changing epidemiology and the impact of highly active antiretroviral therapy. **19**, 14–19.
13. Cheung, M. C., Pantanowitz, L., and Dezube, B. J. (2005) AIDS-related malignancies: emerging challenges in the era of highly active antiretroviral therapy. **10**, 412–426.
14. Beral, V., Peterman, T. A., Berkelman, R. L., and Jaffe, H. W. (1990) Kaposi's sarcoma

among persons with aids: a sexually transmitted infection? **335**, 123–128.

15. Archibald, C. P., Schechter, M. T., Le, T. N., Craib, K. J., Montaner, J. S., and O'shaughnessy, M. V. (1992) Evidence For a sexually transmitted cofactor for AIDS-related Kaposi's sarcoma in a cohort of homosexual men. **3**, 203–209.
16. Vogel, J., Hinrichs, S. H., Reynolds, R. K., Luciw, P. A., and Jay, G. (1988) The HIV Tat gene induces dermal lesions resembling Kaposi's sarcoma in transgenic mice. **335**, 606–611.
17. Ensoli, B., Salahuddin, S. Z., and Gallo, R. C. (1989) AIDS-associated Kaposi's sarcoma: a molecular model for its pathogenesis. **1**, 93–96.
18. Chang, Y., Cesarman, E., Pessin, M. S., Lee, F., Culpepper, J., Knowles, D. M., and Moore, P. S. (1994) Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. **266**, 1865–1869.
19. Russo, J. J., Bohenzky, R. A., Chien, M. C., Chen, J., Yan, M., Maddalena, D., Parry, J. P., Peruzzi, D., Edelman, I. S., Chang, Y., and Moore, P. S. (1996) Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus (HHV8). **93**, 14862–14867.
20. Moore, P. S. and Chang, Y. (1995) Detection of herpesvirus-like DNA sequences in Kaposi's sarcoma in patients with and without HIV infection. **332**, 1181–1185.
21. Cesarman, E., Chang, Y., Moore, P. S., Said, J. W., and Knowles, D. M. (1995) Kaposi's sarcoma-associated herpesvirus-like DNA sequences in AIDS-related body-cavity-based lymphomas. **332**, 1186–1191.
22. Soulier, J., Grollet, L., Oksenhendler, E., Cacoub, P., Cazals-Hatem, D., Babinet, P., D'agay, M. F., Clauvel, J. P., Raphael, M., and Degos, L. (1995) Kaposi's sarcoma-associated herpesvirus-like DNA sequences in multicentric Castleman's disease. **86**, 1276–1280.
23. Codish, S., Abu-Shakra, M., Ariad, S., Zirkin, H. J., Yermiyahu, T., Dupin, N., Boshoff, C., and Sukenik, S. (2000) Manifestations of three HHV-8-related diseases in an HIV-negative patient: immunoblastic variant multicentric Castleman's disease, primary effusion lymphoma, and Kaposi's sarcoma. **65**, 310–314.
24. Ascoli, V., Signoretti, S., Onetti-Muda, A., Pescarmona, E., Della-Rocca, C., Nardi, F., Mastroianni, C. M., Gastaldi, R., Pistilli, A., Gaidano, G., Carbone, A., and Lo-Coco, F. (2001) Primary effusion lymphoma in HIV-infected patients with multicentric Castleman's disease. **193**, 200–209.
25. Sukpanichnant, S., Sivayathorn, A., Visudhiphan, S., and Ngowthammatas, W. (2002) Multicentric Castleman's disease, non-Hodgkin's lymphoma, and Kaposi's sarcoma: a rare simultaneous occurrence. **20**, 127–133.
26. Kasolo, F. C., Mpabalwani, E., and Gompels, U. A. (1997) Infection with AIDS-related herpesviruses in human immunodeficiency virus-negative infants and endemic childhood Kaposi's sarcoma in Africa. **78** (4), 847–855.
27. Sarmati, L. (2004) HHV-8 infection in African children. **11**, 50–53.
28. Andreoni, M., El-Sawaf, G., Rezza, G., Ensoli, B., Nicastri, E., Ventura, L., Ercoli, L., Sarmati, L., and Rocchi, G. (1999) High seroprevalence of antibodies to human herpesvirus-8 in Egyptian children: evidence of nonsexual transmission. **91**, 465–469.
29. Wang, Q. J., Jenkins, F. J., Jacobson, L. P., Kingsley, L. A., Day, R. D., Zhang, Z. W., Meng, Y. X., Pellett, P. E., Kousoulas, K. G., Baghian, A., and Rinaldo, C. R., Jr. (2001) Primary human herpesvirus 8 infection generates a broadly specific cd8(+) T-cell response to viral lytic cycle proteins. **97**, 2366–2373.
30. Luppi, M., Barozzi, P., Schulz, T. F., Setti, G., Staskus, K., Trovato, R., Narni, F., Donelli, A., Maiorana, A., Marasca, R., Sandrini, S., and Torelli, G. (2000) Bone marrow failure associated with human herpesvirus 8 infection after transplantation. **343**, 1378–1385.
31. Sarid, R., Pizov, G., Rubinger, D., Backenroth, R., Friedlaender, M. M., Schwartz, F., and Wolf, D. G. (2001) Detection of human herpesvirus-8 DNA in kidney allografts prior to the development of Kaposi's sarcoma. **32**, 1502–1505.
32. Marcelin, A. G., Roque-Afonso, A. M., Hurtova, M., Dupin, N., Tulliez, M., Sebagh, M., Arkoub, Z. A., Guettier, C., Samuel, D., Calvez, V., and Dussaix, E. (2004) Fatal disseminated Kaposi's sarcoma following human herpesvirus 8 primary infections in liver-transplant recipients. **10,** 295–300.
33. Renne, R., Zhong, W., Herndier, B., Mcgrath, M., Abbey, N., Kedes, D., and Ganem, D. (1996) Lytic growth of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) in culture. **2**, 342–346.
34. Sarid, R., Flore, O., Bohenzky, R. A., Chang, Y., and Moore, P. S. (1998) Transcription

mapping of the Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) genome in a body cavity-based lymphoma cell line (Bc-1). **72**, 1005–1012.

35. Krishnan, H. H., Naranatt, P.P., Smith, M. S., Zeng, L., Bloomer, C., and Chandran, B. (2004) Concurrent expression of latent and a limited number of lytic genes with immune modulation and antiapoptotic function by Kaposi's sarcoma-associated herpesvirus early during infection of primary endothelial and fibroblast cells and subsequent decline of lytic gene expression. **78**, 3601–3620.
36. Grundhoff, A. and Ganem, D. (2004) Inefficient establishment of KSHV latency suggests an additional role for continued lytic replication in Kaposi sarcoma pathogenesis. **113**, 124–136.
37. Chen, J., Ueda, K., Sakakibara, S., Okuno, T., Parravicini, C., Corbellino, M., and Yamanishi, K. (2001) Activation of latent Kaposi's sarcoma-associated herpesvirus by demethylation of the promoter of the lytic transactivator. **98**, 4119–4124.
38. Chang, J., Renne, R., Dittmer, D., and Ganem, D. (2000) Inflammatory cytokines and the reactivation of Kaposi's sarcoma-associated herpesvirus lytic replication. **266**, 17–25.
39. Davis, D. A., Rinderknecht, A. S., Zoeteweij, J. P., Aoki, Y., Read-Connole, E. L., Tosato, G., Blauvelt, A., and Yarchoan, R. (2001) Hypoxia induces lytic replication of Kaposi sarcoma-associated herpesvirus. **97**, 3244–3250.
40. Deutsch, E., Cohen, A., Kazimirsky, G., Dovrat, S., Rubinfeld, H., Brodie, C., and Sarid, R. (2004) Role of protein kinase c delta in reactivation of Kaposi's sarcoma-associated herpesvirus. **78**, 10187–10192.
41. Sgarbanti, M., Arguello, M., Tenoever, B. R., Battistini, A., Lin, R., and Hiscott, J. (2004) A requirement for NF-kappaB induction in the production of replication-competent HHV-8 virions. **23**, 5770–5780.
42. Mcallister, S. C., Hansen, S. G., Messaoudi, I., Nikolich-Zugich, J., and Moses, A. V. (2005) Increased efficiency of phorbol ester-induced lytic reactivation of Kaposi's sarcoma-associated herpesvirus during S phase. **79**, 2626–2630.
43. Brown, H. J., Mcbride, W. H., Zack, J. A., and Sun, R. (2005) Prostratin and bortezomib are novel inducers of latent Kaposi's sarcoma-associated herpesvirus. **10**, 745–751.
44. Chang, M., Brown, H. J., Collado-Hidalgo, A., Arevalo, J. M., Galic, Z., Symensma, T. L., Tanaka, L., Deng, H., Zack, J. A., Sun, R., and Cole, S. W. (2005) Beta-adrenoreceptors reactivate Kaposi's sarcoma-associated herpesvirus lytic replication via PKA-dependent control of viral RTA. **79**, 13538–13547.
45. Morris, T. L., Arnold, R. R., and Webster-Cyriaque, J. (2007) Signaling cascades triggered by bacterial metabolic end-products during KSHV reactivation. J. Virol.
46. Staudt, M. R. and Dittmer, D. P. (2007) The RTA/ORF50 transactivator proteins of the gamma-Herpesviridae. **312**, 71–100.
47. Whitby, D., Marshall, V. A., Bagni, R. K., Miley, W. J., Mccloud, T. G., Hines-Boykin, R., Goedert, J. J., Conde, B. A., Nagashima, K., Mikovits, J., Dittmer, D. P., and Newman, D. J. (2007) Reactivation of Kaposi's sarcoma-associated herpesvirus by natural products from Kaposi's sarcoma endemic regions. **120**, 321–328.
48. Aoki, Y. and Tosato, G. (2007) Interactions between HIV-1 Tat and KSHV. **312**, 309–326.
49. Parravicini, C., Olsen, S. J., Capra, M., Poli, F., Sirchia, G., Gao, S. J., Berti, E., Nocera,A., Rossi, E., Bestetti, G., Pizzuto, M., Galli, M., Moroni, M., Moore, P. S., and Corbellino, M. (1997) Risk of Kaposi's sarcoma-associated herpes virus transmission from donor allografts among Italian posttransplant Kaposi's sarcoma patients. **90**, 2826–2829.
50. Campbell, T. B., Borok, M., Gwanzura, L., Mawhinney, S., White, I. E., Ndemera, B., Gudza, I., Fitzpatrick, L., and Schooley, R.T. (2000) Relationship of human herpesvirus 8 peripheral blood virus load and Kaposi's sarcoma clinical stage. **14**, 2109–2116.
51. Quinlivan, E. B., Zhang, C., Stewart, P. W., Komoltri, C., Davis, M. G., and Wehbie, R. S. (2002) Elevated virus loads of Kaposi's sarcoma-associated human herpesvirus 8 predict Kaposi's sarcoma disease progression, but elevated levels of human immunodeficiency virus type 1 do not. **185**, 1736–1744.
52. Tedeschi, R., Enbom, M., Bidoli, E., Linde, A., De Paoli, P., and Dillner, J. (2001) Viral load of human herpesvirus 8 in peripheral blood of human immunodeficiency virus-infected patients with Kaposi's sarcoma. **39**, 4269–4273.
53. Polstra, A. M., Cornelissen, M., Goudsmit, J., and Van Der Kuyl, A. C. (2004) Retrospective, longitudinal analysis of serum human herpesvirus-8 viral DNA load in aids-related

Kaposi's sarcoma patients before and after diagnosis. **74**, 390–396.

54. Pellet, C., Chevret, S., Frances, C., Euvrard, S., Hurault, M., Legendre, C., Dalac, S., Farge, D., Antoine, C., Hiesse, C., Peraldi, M.N., Lang, P., Samuel, D., Calmus, Y., Agbalika, F., Morel, P., Calvo, F., and Lebbe, C. (2002) Prognostic value of quantitative Kaposi sarcoma--associated herpesvirus load in posttransplantation Kaposi sarcoma. **186**, 110–113.
55. Gessain, A. and Duprez, R. (2005) Spindle cells and their role in Kaposi's sarcoma. **37**, 2457–2465.
56. Montaner, S., Sodhi, A., Molinolo, A., Bugge, T. H., Sawai, E. T., He, Y., Li, Y., Ray, P. E., and Gutkind, J. S. (2003) Endothelial infection with KSHV genes in vivo reveals that vgpcr initiates Kaposi's sarcomagenesis and can promote the tumorigenic potential of viral latent genes. **3**, 23–36.
57. Guo, H. G., Sadowska, M., Reid, W., Tschachler, E., Hayward, G., and Reitz, M. (2003) Kaposi's sarcoma-like tumors in a human herpesvirus 8 orf74 transgenic mouse. **77**, 2631–2639.
58. Grisotto, M. G., Garin, A., Martin, A. P., Jensen, K. K., Chan, P., Sealfon, S. C., and Lira, S. A. (2006) The human herpesvirus 8 chemokine receptor vgpcr triggers autonomous proliferation of endothelial cells. **116**, 1264–1273.
59. Mutlu, A. D., Cavallin, L. E., Vincent, L., Chiozzini, C., Eroles, P., Duran, E. M., Asgari, Z., Hooper, A. T., La Perle, K. M., Hilsher, C., Gao, S. J., Dittmer, D. P., Rafii, S., snd Mesri, E. A. (2007) In vivo-restricted and reversible malignancy induced by human herpesvirus-8 KSHV: a cell and animal model of virally induced Kaposi's sarcoma. **11**, 245–258.
60. Brown, E. E., Whitby, D., Vitale, F., Marshall, V., Mbisa, G., Gamache, C., Lauria, C., Alberg, A. J., Serraino, D., Cordiali-Fei, P., Messina, A., and Goedert, J. J. (2006) Virologic, hematologic, and immunologic risk factors for classic Kaposi sarcoma. **107**, 2282–2290.
61. Hengge, U. R., Ruzicka, T., Tyring, S. K., Stuschke, M., Roggendorf, M., Schwartz, R. A., and Seeber, S. (2002) Update on Kaposi's sarcoma and other hhv8 associated diseases. Part 2: pathogenesis, Castleman's disease, and pleural effusion lymphoma. **2**, 344–352.
62. Cesarman, E. and Mesri, E. A. (2007) Kaposi sarcoma-associated herpesvirus and other viruses in human lymphomagenesis. **312**, 263–87., 263–287.
63. Casper, C. (2005) The aetiology and management of Castleman disease at 50 years: translating pathophysiology to patient care. **129**, 3–17.
64. Whitby, D., Howard, M. R., Tenant-Flowers, M., Brink, N. S., Copas, A., Boshoff, C., Hatzioannou, T., Suggett, F. E., Aldam, D. M., and Denton, A. S. (1995) Detection of Kaposi sarcoma associated herpesvirus in peripheral blood of HIV-infected individuals and progression to Kaposi's sarcoma. **346**, 799–802.
65. Widmer, I. C., Erb, P., Grob, H., Itin, P., Baumann, M., Stalder, A., Weber, R., and Cathomas, G. (2006) Human herpesvirus 8 oral shedding in HIV-infected men with and without Kaposi sarcoma. **42**, 420–425.
66. Boldogh, I., Szaniszlo, P., Bresnahan, W. A., Flaitz, C. M., Nichols, M. C., and Albrecht, T. (1996) Kaposi's Sarcoma herpesvirus-like DNA sequences in the saliva of individuals infected with human immunodeficiency virus. **23**, 406–407.
67. Blackbourn, D. J., Lennette, E. T., Ambroziak, J., Mourich, D. V., and Levy, J. A. (1998) Human herpesvirus 8 detection in nasal secretions and saliva. **177**, 213–216.
68. Pauk, J., Huang, M. L., Brodie, S. J., Wald, A., Koelle, D. M., Schacker, T., Celum, C., Selke, S., and Corey, L. (2000) Mucosal shedding of human herpesvirus 8 in men. **343**, 1369–1377.
69. Pellett, P. E., Spira, T. J., Bagasra, O., Boshoff, C., Corey, L., De Lellis, L., Huang, M. L., Lin, J. C., Matthews, S., Monini, P., Rimessi, P., Sosa, C., Wood, C., and Stewart, J. A. (1999) Multicenter comparison of PCR assays for detection of human herpesvirus 8 DNA in semen. **37**, 1298–1301.
70. Diamond, C., Brodie, S. J., Krieger, J. N., Huang, M. L., Koelle, D. M., Diem, K., Muthui, D., and Corey, L. (1998) Human herpesvirus 8 in the prostate glands of men with Kaposi's sarcoma. **72**, 6223–6227.
71. Cerimele, F., Curreli, F., Ely, S., Friedman-Kien, A. E., Cesarman, E., and Flore, O. (2001) Kaposi's sarcoma-associated herpesvirus can productively infect primary human keratinocytes and alter their growth properties. **75**, 2435–2443.
72. Akula, S. M., Wang, F. Z., Vieira, J., and Chandran, B. (2001) Human herpesvirus 8 interaction with target cells involves heparan sulfate. **282**, 245–255.
73. Krishnan, H. H., Sharma-Walia, N., Zeng, L., Gao, S. J., and Chandran, B. (2005) Envelope glycoprotein gb of Kaposi's

sarcoma-associated herpesvirus is essential for egress from infected cells. **79**, 10952–10967.

74. Martin, J. N., Ganem, D. E., Osmond, D. H., Page-Shafer, K. A., Macrae, D., and Kedes, D. H. (1998) Sexual transmission and the natural history of human herpesvirus 8 infection. **338**, 948–954.
75. Blackbourn, D. J., Osmond, D., Levy, J. A., and Lennette, E. T. (1999) Increased human herpesvirus 8 seroprevalence in young homosexual men who have multiple sex contacts with different partners. **179**, 237–239.
76. Grulich, A. E., Kaldor, J. M., Hendry, O., Luo, K., Bodsworth, N. J., and Cooper, D. A. (1997) Risk of Kaposi's sarcoma and oroanal sexual contact. **145**, 673–679.
77. Smith, N. A., Sabin, C. A., Gopal, R., Bourboulia, D., Labbet, W., Boshoff, C., Barlow, D., Band, B., Peters, B. S., De Ruiter, A., Brown, D. W., Weiss R. A., Best, J. M., and Whitby, D. (1999) Serologic evidence of human herpesvirus 8 transmission by homosexual but not heterosexual sex. **180**, 600–606.
78. Osmond, D. H., Buchbinder, S., Cheng, A., Graves, A., Vittinghoff, E., Cossen, C. K., Forghani, B., and Martin, J. N. (2002) Prevalence of Kaposi sarcoma-associated herpesvirus infection in homosexual men at beginning of and during the HIV epidemic. **287**, 221–225.
79. Mayama, S., Cuevas, L. E., Sheldon, J., Omar, O.H., Smith, D. H., Okong, P., Silvel, B., Hart, C. A., and Schulz, T. F. (1998) Prevalence and transmission of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) in Ugandan children and adolescents. 77, 817–820.
80. Gessain, A., Mauclere, P., Van Beveren, M., Plancoulaine, S., Ayouba, A., Essame-Oyono, J. L., Martin, P. M., and De The, G. (1999) Human herpesvirus 8 primary infection occurs during childhood in Cameroon, Central Africa. **81**, 189–192.
81. Bourboulia, D., Whitby, D., Boshoff, C., Newton, R., Beral, V., Carrara, H., Lane, A., and Sitas, F. (1998) Serologic evidence for mother-to-child transmission of Kaposi sarcoma-associated herpesvirus infection (letter). **280**, 31–32.
82. Plancoulaine, S., Abel, L., Van Beveren, M., Tregouet, D. A., Joubert, M., Tortevoye, P., De The, G., and Gessain, A. (2000) Human herpesvirus 8 transmission from mother to child and between siblings in an endemic population. **356**, 1062–1065.
83. Sarmati, L., Carlo, T., Rossella, S., Montano, M., Adalgisa, P., Rezza, G., and Andreoni, M. (2004) Human herpesvirus-8 infection in pregnancy and labor: lack of evidence of vertical transmission. **72**, 462–466.
84. Angeloni, A., Heston, L., Uccini, S., Sirianni, M. C., Cottoni, F., Masala, M. V., Cerimele, D., Lin, S. F., Sun, R., Rigsby, M., Faggioni, A., and Miller, G. (1998) High prevalence of antibodies to human herpesvirus 8 in relatives of patients with classic Kaposi's sarcoma from Sardinia. **177**, 1715–1718.
85. Guttman-Yassky, E., Kra-Oz, Z., Dubnov, J., Friedman-Birnbaum, R., Segal, I., Zaltzman, N., Roth, T., Schwartz, F., Linn, S., Rozenman, D., David, M., Silbermann, M., Barchana, M., Bergman, R., and Sarid, R. (2005) Infection with Kaposi's sarcoma-associated herpesvirus among families of patients with classic Kaposi's sarcoma. **141**, 1429–1434.
86. Davidovici, B., Karakis, I., Bourboulia, D., Ariad, S., Zong, J. C., Benharroch, D., Dupin,N., Weiss, R., Hayward, G., Sarov, B., and Boshoff, C. (2001) Seroepidemiology and molecular epidemiology of Kaposi's sarcoma-associated herpesvirus among Jewish population groups in Israel. **93**, 194–202.
87. Plancoulaine, S., Abel, L., Tregouet, D., Duprez, R., Van Beveren, M., Tortevoye, P., Froment, A., and Gessain, A. (2004) Respective roles of serological status and blood specific antihuman herpesvirus 8 antibody levels in human herpesvirus 8 intrafamilial transmission in a highly endemic area. **64**, 8782–8787.
88. Mbulaiteye, S. M., Biggar, R. J., Bakaki, P. M., Pfeiffer, R. M., Whitby, D., Owor, A. M., Katongole-Mbidde, E., Goedert, J. J., Ndugwa, C. M., and Engels, E. A. (2003) Human herpesvirus 8 infection and transfusion history in children with sickle-cell disease in Uganda. **95**, 1330–1335.
89. Hladik, W., Dollard, S. C., Mermin, J., Fowlkes, A. L., Downing, R., Amin, M. M., Banage, F., Nzaro, E., Kataaha, P., Dondero, T. J., Pellett, P.E., and Lackritz, E. M. (2006) Transmission of human herpesvirus 8 by blood transfusion. **355**, 1331–1338.
90. Barozzi, P., Luppi, M., Facchetti, F., Mecucci, C., Alu, M., Sarid, R., Rasini, V., Ravazzini,L., Rossi, E., Festa, S., Crescenzi, B., Wolf, D. G., Schulz, T. F., and Torelli, G. (2003) Post-transplant Kaposi sarcoma originates from the seeding of donor-derived progenitors. 9, 554–561.
91. Parkin, D. M., Whelan, S. L., Ferlay, J., Raymond, L., and Young, J. (2003) Cancer

Incidence In Five Continents, Vol. VIII. IARC Scientific Publications, Lyon, France.

92. Gao, S. J., Kingsley, L., Li, M., Zheng, W., Parravicini, C., Ziegler, J., Newton, R., Rinaldo, C. R., Saah, A., Phair, J., Detels, R., Chang, Y., and Moore, P. S. (1996) KSHV antibodies among Americans, Italians and Ugandans with and without Kaposi's sarcoma. **2**, 925–928.
93. Gao, S. J., Kingsley, L., Hoover, D. R., Spira, T. J., Rinaldo, C. R., Saah, A., Phair, J., Detels, R., Parry, P., Chang, Y., and Moore, P. S. (1996) Seroconversion to antibodies against Kaposi's sarcoma-associated herpesvirus-related latent nuclear antigens before the development of Kaposi's sarcoma. **335**, 233–241.
94. Kedes, D. H., Operskalski, E., Busch, M., Kohn, R., Flood, J., and Ganem, D. (1996) The seroepidemiology of human herpesvirus 8 (Kaposi's sarcoma-associated herpesvirus): distribution of infection in KS risk groups and evidence for sexual transmission. **2**, 918–924.
95. Simpson, G. R., Schulz, T. F., Whitby, D., Cook, P. M., Boshoff, C., Rainbow, L., Howard, M. R., Gao, S. J., Bohenzky, R. A., Simmonds, P., Lee, C., De Ruiter, A., Hatzakis, A., Tedder, R. S., Weller, I. V., Weiss, R. A., and Moore, P. S. (1996) Prevalence of Kaposi's sarcoma associated herpesvirus infection measured by antibodies to recombinant capsid protein and latent immunofluorescence antigen. **348**, 1133–1138.
96. Laney, A. S., Peters, J. S., Manzi, S. M., Kingsley, L. A., Chang, Y., and Moore, P. S. (2006) Use of a multiantigen detection algorithm for diagnosis of Kaposi's sarcoma-associated herpesvirus infection. **44**, 3734–3741.
97. Dukers, N. H. and Rezza, G. (2003) Human herpesvirus 8 epidemiology: what we do and do not know. **17**, 1717–1730.
98. Mohanna, S., Maco, V., Bravo, F., and Gotuzzo, E. (2005) Epidemiology and clinical characteristics of classic Kaposi's sarcoma, seroprevalence, and variants of human herpesvirus 8 in South America: a critical review of an old disease. **9**, 239–250.
99. Dedicoat, M. and Newton, R. (2003) Review of the distribution of Kaposi's sarcoma-associated herpesvirus (KSHV) in Africa in relation to the incidence of Kaposi's sarcoma. 88, 1–3.
100. Plancoulaine, S., Gessain, A., Van Beveren, M., Tortevoye, P., and Abel, L. (2003) Evidence for a recessive major gene predisposing to human herpesvirus 8 (HHV-8) infection in a population in which HHV-8 is endemic. **187**, 1944–1950.
101. Ariyoshi, K., Schim, V. D. L., Cook, P., Whitby, D., Corrah, T., Jaffar, S., Cham, F., Sabally, S., O'donovan, D., Weiss, R. A., Schulz, T. F., and Whittle, H. (1998) Kaposi's sarcoma in The Gambia, West Africa is less frequent in human immunodeficiency virus type 2 than in human immunodeficiency virus type 1 infection despite a high prevalence of human herpesvirus 8. **1**, 193–199.
102. He, F., Wang, X., He, B., Feng, Z., Lu, X., Zhang, Y., Zhao, S., Lin, R., Hui, Y., Bao, Y., Zhang, Z., and Wen, H. (2007) Human herpesvirus 8: serovprevalence and correlates in tumor patients from Xinjiang, China. **79**, 161–166.
103. Biggar, R. J., Whitby, D., Marshall, V., Linhares, A. C., and Black, F. (2000) Human herpesvirus 8 in Brazilian Amerindians: a hyperendemic population with a new subtype. **181**, 1562–1568.
104. Grossman, Z., Iscovich, J., Schwartz, F., Azizi, E., Klepfish, A., Schattner, A., and Sarid, R. (2002) Absence of Kaposi sarcoma among Ethiopian immigrants to Israel despite high seroprevalence of human herpesvirus 8. 77, 905–909.
105. Cottoni, E., Masia, I. M., Masala, M. V., Mulargia, M., and Contu, L. (1996) Familial Kaposi's sarcoma: case reports and review of the literature. **76**, 59–61.
106. Guttman-Yassky, E., Cohen, A., Kra-Oz, Z., Friedman-Birnbaum, R., Sprecher, E., Zaltzman, N., Friedman, E., Silbermann, M., Rubin, D., Linn, S., Whitby, D., Gideoni, O., Pollack, S., Bergman, R., and Sarid, R. (2004) Familial clustering of classic Kaposi sarcoma. **189**, 2023–2026.
107. Hayward, G. S. and Zong, J. C. (2007) Modern evolutionary history of the human KSHV genome. **312**, 1–42.
108. Sarid, R., Olsen, S. J., and Moore, P. S (1999) Kaposi's sarcoma-associated herpesvirus epidemiology, virology, and molecular biology. **52**, 139–232.
109. Vitale, F., Briffa, D. V., Whitby, D., Maida, I., Grochowska, A., Levin, A., Romano, N., and Goedert, J. J. (2001) Kaposi's Sarcoma herpes virus and Kaposi's sarcoma in the elderly populations of 3 Mediterranean islands. **91**, 588–591.
110. Woodle, E. S., Hanaway, M., Buell, J., Gross, T., First, M. R., Trofe, J., and Beebe, T. (2001) Kaposi sarcoma: an analysis of the us and international experiences from The Israel Penn International Transplant Tumor Registry. **33**, 3660–3661.

111. Renwick, N., Halaby, T., Weverling, G. J., Dukers, N. H., Simpson, G. R., Coutinho, R. A., Lange, J. M., Schulz, T. F., and Goudsmit, J. (1998) Seroconversion for human herpesvirus 8 during HIV infection is highly predictive of Kaposi's sarcoma. **12**, 2481–2488.

112. Dorak, M. T., Yee, L. J., Tang, J., Shao, W., Lobashevsky ,E. S., Jacobson, L. P., and Kaslow, R. A. (2005) Hla-B, -Drb1/3/4/5, and -Dqb1 gene polymorphisms in human immunodeficiency virus-related Kaposi's sarcoma. **76**, 302–310.

113. Gaya, A., Esteve, A., Casabona, J., Mccarthy, J. J., Martorell, J., Schulz, T. F., and Whitby, D. (2004) Amino acid residue at position 13 in HLA-dr beta chain plays a critical role in the development of Kaposi's sarcoma in AIDS patients. **18**, 199–204.

114. Cottoni, F., Masala, M. V., Santarelli, R., Carcassi, C., Uccini, S., Montesu, M. A., Satta, R., Angeloni, A., Faggioni, A., and Cerimele, D. (2004) Susceptibility to human herpesvirus-8 infection in a healthy population from Sardinia is not directly correlated with the expression of HLA-dr alleles. **151**, 247–249.

115. Foster ,C. B., Lehrnbecher, T., Samuels, S., Stein, S., Mol, F., Metcalf, J. A., Wyvill, K., Steinberg, S. M., Kovacs, J., Blauvelt, A., Yarchoan, R., and Chanock, S. J. (2000) An IL6 promoter polymorphism is associated with a lifetime risk of development of Kaposi sarcoma in men infected with human immunodeficiency virus. **96**, 2562–2567.

116. Lehrnbecher, T. L., Foster, C. B., Zhu, S., Venzon, D., Steinberg, S. M., Wyvill, K., Metcalf, J. A., Cohen, S. S., Kovacs, J., Yarchoan,R., Blauvelt, A., and Chanock, S. J. (2000) Variant genotypes of fcgammariiia influence the development of Kaposi's sarcoma in HIV-infected men. **95**, 2386–2390.

117. Brown, E. E., Fallin, D., Ruczinski, I., Hutchinson, A., Staats, B., Vitale, F., Lauria, C., Serraino, D., Rezza, G., Mbisa, G., Whitby, D., Messina, A., Goedert, J. J., and Chanock, S. J. (2006) Associations of classic Kaposi sarcoma with common variants in genes that modulate host immunity. **15**, 926–934.

118. Simonart, T. (2006) Role of environmental factors in the pathogenesis of classic and African-endemic Kaposi sarcoma. **244**, 1–7.

119. Montella, M., Serraino, D., Crispo, A., Rezza, G., Carbone, S., and Tamburini, M. (2000) Is volcanic soil a cofactor for classic Kaposi's sarcoma? **16**, 1185–1186.

120. Nawar, E., Mbulaiteye, S. M., Gallant, J. E., Wohl, D. A., Ardini, M., Hendershot, T., Goedert, J. J., and Rabkin, C. S. (2005) Risk factors for Kaposi's sarcoma among hhv-8 seropositive homosexual men with AIDS. **115**, 296–300.

121. Goedert, J. J., Vitale, F., Lauria, C., Serraino, D., Tamburini, M., Montella, M., Messina, A., Brown, E. E., Rezza, G., Gafa, L., and Romano, N. (2002) Risk factors for classical Kaposi's sarcoma. **94**, 1712–1718.

122. Guttman-Yassky, E., Dubnov, J., Kra-Oz, Z., Friedman-Birnbaum, R., Silbermann, M., Barchana, M., Bergman, R., and Sarid, R. (2006) Classic Kaposi sarcoma which KSHV-seropositive individuals are at risk? **106**, 413–419.

123. Pfeffer, S., Sewer, A., Lagos-Quintana, M., Sheridan, R., Sander, C., Grasser, F. A., Van Dyk, L. F., Ho, C. K., Shuman, S., Chien, M., Russo, J. J., Ju, J., Randall, G., Lindenbach, B. D., Rice, C. M., Simon, V., Ho, D. D., Zavolan, M., and Tuschl,T. (2005) Identification of micrornas of the herpesvirus family. **2**, 269–276.

124. Cai, X., Lu, S., Zhang, Z., Gonzalez, C. M., Damania, B., and Cullen, B. R. (2005) Kaposi's sarcoma-associated herpesvirus expresses an array of viral microrna in latently infected cells. **102**, 5570–5575.

125. Samols, M. A., Hu, J., Skalsky, R. L., and Renne, R. (2005) Cloning and identification of a microrna cluster within the latency-associated region of Kaposi's sarcoma-associated herpesvirus. **79**, 9301–9305.

126. Grundhoff, A., Sullivan, C. S., and Ganem, D. (2006) A combined computational and microarray-based approach identifies novel microrna encoded by human gamma-herpesviruses. **12**, 733–750.

127. Moore, P. S. and Chang, Y. (2003) Kaposi's sarcoma-associated herpesvirus immunoevasion and tumorigenesis: two sides of the same coin? **57**, 609–39., 609–639.

128. Schulz, T. F. (2006) The pleiotropic effects of Kaposi's sarcoma herpesvirus. 208, 187–198.

129. Jarviluoma, A. and Ojala, P. M. (2006) Cell signaling pathways engaged by KSHV. **1766**, 140–158.

130. Rezaee, S. A., Cunningham, C., Davison, A. J., and Blackbourn, D. J. (2006) Kaposi's sarcoma-associated herpesvirus immune modulation: an overview. **87**, 1781–1804.

131. Sodhi,A., Montaner,S., And Gutkind,J.S. (2004) Does dysregulated expression of a deregulated viral gpcr trigger Kaposi's sarcomagenesis? **18**, 422–427.

132. Boshoff,C. (2003) Kaposi Virus Scores Cancer Coup. **9**, 261–262.

Chapter 20

Cancer Cohort Consortium Approach: Cancer Epidemiology in Immunosuppressed Groups

Diego Serraino, Pierluca Piselli for the Study Group[1]

Summary

Nearly 40 years have passed since the publication of the first report showing higher cancer risks in recipients of organ transplants. Thereafter, studies carried out in immunosuppressed persons have greatly expanded our knowledge on the spectrum of cancers associated with infections. Clinical investigations following the expanding practice of organ transplantation, and since the 1980s, studies on human immunodeficiency virus (HIV)-infected persons have thereafter confirmed and extended these early observations. The comparison of the spectrum of cancers seen in excess in these two groups of populations offers an original viewpoint into the association of immunosuppression and cancer. Combining longitudinal data from different cohorts of HIV-infected persons and of transplant recipients represents a further tool for better quantifying the spectrum of cancers associated with immunodepression. In this chapter, the use of this methodologic approach in southern Europe is illustrated and limitations are discussed.

Key words: Cancer risk, HIV-infection, immunosuppression, multi-cohort, transplant recipients, southern Europe.

1. Introduction

Doll and Kinlen firstly noted, in 1970, an increased frequency of certain cancers in people treated with antirejection drugs after organ transplantation. The same malignancies, namely, non-Hodgkin lymphoma (NHL), Kaposi's sarcoma (KS), nonmelanoma skin cancers, anogenital cancers, and Hodgkin

[1] For the Immunosuppression and Cancer Study Group, members are listed at the end of this chapter.

Mukesh Verma (ed.), *Methods in Molecular Biology, Cancer Epidemiology, vol. 471*
© 2009 Humana Press, a part of Springer Science + Business Media, Totowa, NJ
Book doi: 10.1007/978-1-59745-416-2

lymphoma (HL), turned out to be the commonest manifestations of human immunodeficiency virus (HIV) infection and AIDS *(1–3)*. Other cancers related to lifestyle risk factors (e.g., such as lung cancer and smoking, liver cancer and alcohol consumption) were also found with increasing frequency in HIV-infected persons or in transplant recipients, but the role of immunosuppression has not been well quantified yet *(4,5)*.

Several types of epidemiologic investigations, including record linkage of population-based registries (cancer, HIV/AIDS, recipients of organ transplants), follow-up and case-control studies, have contributed to increase our knowledge on the spectrum of cancer related to immunosuppression *(5)*. In southern Europe, where the range of cancers seen in HIV-infected persons has been deeply investigated, the frequency of organ transplants has doubled in the last decade, but few studies have quantified the cancer risk of organ transplant recipients *(6–8)*. Studies of cancer in different immunodepressed populations have the potential to help delineate which cancers are in fact linked to immunosuppression, given that lifestyle-related risk factors substantially differ in these population groups *(5)*.

This chapter illustrate and discusses the methodology used and the results obtained by combining, in southern Europe, two cohorts of HIV-infected individuals and five cohorts of recipients of organ transplants.

2. Methods

2.1. Data Sources

2.1.1. HIV-Infected Individuals

Follow-up data were collected from a cohort of seroprevalent persons enrolled in Nice, France, through the nationwide database Dossier Médical Informatique-2 (DMI-2) and from an Italian cohort of persons with known date of HIV seroconversion (the Italian HIV Seroconversion Study [ISS]) (see general characteristics in **Table 20.1**). Epidemiologic and clinical information on all HIV-infected individuals who have access to hospital care in France are recorded through the DMI-2. For this analysis, we used the DMI-2 of Provence-Cote d'Azur region, southern France, regarding 6,072 individuals newly diagnosed with HIV infection between 1988 and 2004. These people underwent medical examination in the HIV clinics located in the study area at enrollment and, on average, every 6 mo. They were followed-up for a median time of 3.6 years (interquartile range [IQR]: 1.5–7.0).

The ISS is an ongoing multicenter cohort investigation of individuals with a known date of seroconversion, enrolled in 18 clinical centers throughout Italy. These people had a documented

Table 20.1
Main characteristics of HIV-infected persons and transplant recipients in Italy and France

Characteristics	HIV-positive persons	Transplant recipients				
		Kidney	Heart	Liver	Lung	All
	No. (%)	No. (%)	No. (%)	No. (%)	No. (%)	No. (%)
Sex						
Men	5 743 (71.1)	1 370 (64.6)	564 (82.8)	228 (70.6)	29 (69.0)	2 191 (69.2)
Women	2 331 (28.9)	750 (35.4)	117 (17.2)	95 (29.4)	13 (31.0)	975 (30.8)
Age at enrollment (years)						
<35	5 473 (67.8)	681 (32.1)	100 (14.7)	32 (9.9)	5 (11.9)	818 (25.8)
35–49	2 120 (26.3)	840 (39.6)	197 (28.9)	113 (35.0)	18 (42.9)	1 168 (36.9)
≥50	481 (5.9)	599 (28.3)	384 (56.4)	178 (55.1)	19 (45.2)	1 180 (37.3)
Calendar period at enrollment						
<1990	3 068 (38.0)	706 (33.3)	188 (27.6)	0 (0.0)	0 (0.0)	894 (28.2)
1991–1995	2 946 (36.5)	366 (17.3)	185 (27.2)	77 (23.8)	18 (42.9)	646 (20.4)
≥1996	2 060 (25.5)	1 048 (49.4)	308 (45.2)	246 (76.2)	24 (57.1)	1 626 (51.4)
Total	8 074 (100)	2 120 (100)	681 (100)	323 (100)	42 (100)	3 166 (100)
PYs of follow-up	44 400	16 879	4 840	1 850	97	23 665
Median years of follow-up (IQR)	3.4 (1.5–11.7)	6.8 (2.7–11.6)	6.6 (2.8–10.7)	6.1 (1.8–8.9)	1.7 (0.4–3.8)	6.6 (2.5–10.8)

HIV-seronegative test, followed by a positive test (with a maximum accepted lag-time between the two tests of 3 years). The midpoint between the two tests was used to estimate the seroconversion time point. Between 1985 and 2005, 2002 individuals were enrolled, and they were followed-up for a median time of 8.1 years (IQR: 4.8–11.7).

Overall, this statistical analysis concerned 8,074 HIV-positive persons followed up for 44 400 person-years (PYs) (**Table 20.1**). HIV-infected people had a median age of 31.3 years (interquartile range [IQR]: 26.7–37.1): 28.9% of them was represented by women. With regard to period at enrollment, a quarter of HIV-infected people was enrolled after 1995, i.e., after the introduction of highly active antiretroviral therapies (HAART) (**Table 20.1**).

2.1.2. Transplant Recipients

Information was collected on 3,166 people who resided throughout Italy and who underwent organ transplantation in northern (Milan, Padua, Pavia, and Udine) or central Italy (Rome) (**Table 20.1**). Of these, 2,120 (64.6% men; median age 42.4 years) received renal transplant at Niguarda Hospital, Milan, and at Policlinico Gemelli Hospital, Rome; 681 (82.8% men; median age 52.1 years) received heart transplant at Policlinico S. Matteo, Pavia; 323 (70.6% men; median age 51.3 years) received liver transplant at the University Medical School of Padua and at the University Medical School of Udine; and 42 (69.0% men; median age 48.8 years) underwent lung transplant at Niguarda Hospital, Milan (**Table 20.1**).

2.1.3. Exclusion Criteria

To avoid inclusion of prevalent cancers, we excluded from the present analysis individuals with lifetime history of any cancer; individuals who developed cancer within 30 days after transplant or after enrollment in the HIV cohorts; individuals of both groups who had died within 30 days after transplant or after enrollment in the HIV cohorts.

2.1.4. Baseline Variables

A minimum data set of variables common to both HIV-infected persons was 1) extracted from already existing databases made up for clinical purposes, or 2) collected ad hoc for the aim of this investigation. Collected information included sociodemographic characteristics (sex, date and place of birth, place of residence), date of enrollment (i.e., date of first transplant, or date of HIV seroconversion, or date of first HIV-positive test), HIV-transmission category, use of HAART, type of transplant, and antirejection therapies.

2.1.5. Cancer Diagnoses and Vital Status

Follow-up visits were scheduled at least every 6 months for HIV-infected people, and at various intervals, in accordance with immunosuppressive protocols, in transplant recipients. PYs at risk for cancer were computed from 30 days after enrollment (i.e., the date of transplant, or the date of enrollment in the HIV-cohorts) to the date of last follow-up visit, or to the date of cancer diagnosis , or to the date of death. Observed cancers were the incident cases diagnosed in the cohorts during the study period. Cancer diagnoses were recorded during the follow-up visits, histologically

confirmed and coded according to the International Classification of Diseases and Related Health Problems, 10th revision (ICD-10) *(9)*. To avoid follow-up losses, information on cancer and on vital status was actively elicited either from the national AIDS registries, clinical records, or the census bureau of the town of residence. Multiple primary cancers were separately considered in the statistical analysis. Nonmelanoma skin cancers in situ and preneoplastic lesions were not included in the present analysis as 1) information on basal cell carcinoma was not recorded, and 2) the report of squamous cell carcinoma might not have been complete.

2.1.6. Population-Based Cancer Registries

The number of observed incident cancer cases was compared with the expected number. This was computed from sex- and age-specific incidence rates from Italian and French cancer registries for the periods 1987–1991 *(10)* or 1992–1997 *(11)*. For transplant recipients, the expected number of cancers was computed according to corresponding period of diagnosis of observed cancer and area of residence within Italy (i.e., northern, central, and southern Italy). Because most of the cancers associated with HIV infection were very rare in young adults before the AIDS epidemic, the expected number of cancers in HIV-infected persons was computed using only the first period. Because information on town of residence was not available for the totality of Italian and French HIV-positive persons, the combination of all of Italian and all of French cancer registries was used.

2.2. Statistical Analysis

Standardized incidence ratios (SIRs) were computed dividing the number of observed cases by the number of expected ones. Ninety-five percent confidence intervals (CIs) of SIRs were determined using the Poisson distribution *(12)*. Because of the small number of lung transplant recipients and the similarity of immunosuppressive protocols used, SIRs regarding heart and lung transplant recipients were combined. The cumulative probabilities of developing cancer after transplant or, in the ISS, after HIV seroconversion were estimated according to Kaplan–Meyer. The log-rank test was used to compare the probabilities of cancer development within subgroups.

3. Results

In total, 11,240 people (8,074 HIV-infected individuals and 3,166 transplant recipients) were included and followed up for 68 065 PYs (**Table 20.1**). After 5 years since transplantation or HIV seroconversion, the probability of developing any type of

cancer (excluding nonmelanoma skin cancers) was 3.1% among HIV-positive individuals and 4.4% among transplant recipients. After 15 years, the cumulative probabilities were similar in the

Table 20.2
SIRs and 95% CIs for cancers[a] in HIV-positive people and in transplant recipients in Italy and France

	HIV*-positive persons			Transplant recipients		
Cancer site (ICD-10 code)	Cases	SIR[b]	(95% CI)[c]	Cases	SIR[b]	(95% CI)[c]
Head and neck (C 00–14; 30, 32)	10	1.2	(0.6–2.1)	13	4.8	2.6–8.3
Stomach (C 16)	4	1.8	(0.5–4.5)	11	1.9	(0.4–5.6)
Colon (C 18)	1	0.3	(0.0–1.6)	14	1.5	(0.8–2.5)
Anus (C 21)	5	32.5	(10.6–75.9)	0	0.0	(0.0–10.8)
Liver (C 22)	11	9.4	(4.7–16.9)	10	2.6	(1.2–4.8)
Pancreas (C 25)	2	2.3	(0.3–8.3)	6	2.1	(0.8–4.5)
Lung (C 34)	14	1.7	(0.9–2.8)	32	1.7	(1.2–2.4)
Melanoma of skin (C 43)	2	1.2	(0.1–4.2)	6	2.0	(0.7–4.4)
KS (C 46)	317	451	(403–504)	42	93	(67–126)
Breast, female (C 50)	5	0.8	(0.3–1.9)	12	1.1	(0.6–1.9)
Cervix uteri (C 53)	22	14.6	(9.1–22.1)	3	3.2	(0.7–9.5)
Uteri, corpus (C 54)	0	0.0	(0.0–6.8)	8	4.5	(1.9–8.9)
Kidney (C 64)	0	0.0	(0.0–1.7)	13	3.3	(1.7–5.6)
Bladder (C 67)	2	0.7	(0.1–2.4)	5	0.5	(0.2–1.3)
HL (C 81)	18	10.8	(6.4–17.0)	0	0.0	(0.0–3.4)
NHL (C 82–85; 96)	201	62	(54–71.3)	45	9.9	(7.2–14.4)
All cancers	625	9.8	(9.0–10.6)	243	2.2	(1.9–2.5)
All cancers except KS and NHL	107	1.8	(1.5–2.2)	159	1.8	(1.5–2.1)

[a] Includes cancer types with at least five observed cases in one of the two groups.
[b] SIR = Standardized Incidence Ratios
[c] CI = Confidence Intervals

two groups, 13.3% in HIV positives (95% CI =11.2–15.9%) and 14.7% in transplant recipients (95% CI =12.7–17.0%).

SIRs for cancer sites or types with at least five observed cases in either of the two study groups are shown in **Table 20.2**. The SIR for all cancer sites was 9.8 (95% CI = 9.0–10.6) in HIV-infected individuals and 2.2 (95% CI = 1.9–2.5) in transplant recipients. After the exclusion of KS and NHL, the SIRs for all cancers became similar (1.8 in both groups) (**Table 20.2**).

Significantly increased risks were seen, in both groups, for KS (SIR = 451 in HIV-infected people and 93 in transplant recipients), NHL (SIR = 62 and 10), and liver cancer (SIR = 9.4 and 2.6) (**Table 20.2**). An increase of 1.7-fold was noted in the risk for lung cancer in both groups, whereas significantly elevated SIRs for anal and cervical cancers and for Hodgkin lymphoma (HD) were found among HIV-infected people only. Conversely, head and neck cancers (SIR = 4.8) and kidney cancer (SIR = 3.3) were significantly increased only among recipients of organ transplants (**Table 20.2**).

With regard to gender, SIRs for all cancers were similar in transplant recipients (SIR = 2.2 in both sexes), whereas a significant difference was noted between HIV-infected men and HIV-infected women (10.8 and 6.4, respectively). This difference tended to decrease when NHL and KS were excluded from the analysis (**Table 20.3**). In particular, SIRs for KS were higher in women than in men in both groups, although 95% CIs widely overlapped.

4. Comments

The tendency of HIV infection to become a chronic illness allows comparison with other population groups with prolonged deficit of the immune system, namely, organ transplant recipients. In southern Europe, the occurrence of cancer in HIV-infected people has been widely investigated *(4,13)*, and some of the cancers seen in excess (e.g., KS) are frequently found in the general population. In contrast, few large-scale epidemiologic studies have quantified the posttransplant risk of cancer *(6–8)*.

This multi-cohort approach used already available data to simultaneously investigate the risk of cancer in HIV-infected persons and in transplant recipients. The findings were in agreement with those from a meta-analysis of population-based studies *(5)*, and they confirmed the prominent role of oncogenic viruses in the risk of cancers that follow HIV infection and organ transplantation. Findings regarding lung and liver cancers in transplant recipients is southern Europe were less known, and they put new

Table 20.3
SIRs and 95% confidence intervals (CIs) for cancers with ≥10 observed cases in HIV-infected people or in transplant recipients, according to gender. Italy and France

	HIV-infected persons		Transplant recipients	
	Men SIR[a] (95% CI)[b]	Women SIR[a] (95% CI)[b]	Men SIR[a] (95% CI)[b]	Women SIR[a] (95% CI)[b]
Head and neck	1.1 (0.5–2.1)	3.4 (0.1–18.8)	5.2(2.5–9.5)	4.0(0.8–11.8)
Stomach	1.5 (0.3–4.4)	3.7 (0.1–20.5)	2.2(1.1–3.8)	1.8(0.2–6.6)
Colon	0.4 (0.0–2.0)	0.0 (0.0–4.8)	1.7 (0.9–3.0)	0.9 (0.1–3.3)
Liver	9.1 (4.4–16.8)	14.9 (0.4–83)	2.0 (0.8–4.1)	8.3 (1.7–24.2)
Lung	1.6 (0.9–2.8)	3.3 (0.1–18.2)	1.7 (1.2–2.4)	1.5 (0.2–5.6)
KS	443 (395–497)	639 (379–1,009)	83 (60–116)	243 (98–501)
Kidney	0.0 (0.0–1.9)	0.0 (0.0–13.7)	3.6 (1.8–6.2)	1.7 (0.0–9.2)
NHL	60.1 (51–70)	70.4 (51–95)	10.4 (7.3–14.4)	8.2 (3.7–15.5)
All cancers	10.8 (9.9–11.8)	6.4 (5.2–7.8)	2.2 (1.9–2.6)	2.2 (1.5–2.5)
All cancers except KS and NHL	1.5 (1.2–1.9)	2.6 (1.8–3.5)	1.9 (1.6–2.3)	1.5 (1.1–2.0)

[a]SIR = Standardized Incidence Ratios
[b]CI = Confidence Intervals

questions on the role of immunosuppression on the occurrence of cancers strongly related to life style risk factors (e.g., smoking and alcohol consumption).

Although availability of already collected data for clinical purposes is the main advantage of this type of study design, combining heterogeneous data from different sources may pose serious methodologic problems. Some of the main limitations of this study design (i.e., geographical and group heterogeneity, statistical considerations, cancer ascertainment, and computation of expected numbers of cancer cases) deserve consideration for a thorough interpretation of the results.

4.1. Heterogeneity and Statistical Analysis

In this analysis, we pooled data from two HIV cohorts with different characteristics (the French HIV cohort is defined by access

to hospital care, whereas the Italian HIV cohort is defined by a known date of seroconversion) and from transplant cohorts. Members of the Italian HIV cohort were located throughout Italy, whereas the solid organ transplant recipients were mostly from northern or central Italy. Significant differences in the within-country geographical locations of the HIV patients and transplant recipients may make comparisons difficult for those cancers with known environmental risk factors (e.g., cutaneous melanoma and ultraviolet radiation) and those cancers with a known or strongly suspected viral association (e.g,. KS and human herpes virus 8 infection; liver cancer, NHL, and infection with hepatitis C virus) where the viral seroprevalence may differ by region. Moreover, age and sex distribution of HIV-infected individuals and of transplant recipients and life style factors associated with the probability of cancer occurrence differ in these two groups.

With regard to HIV-infected individuals, we arbitrarily assumed that their age and sex distribution and their mode of HIV acquisition were comparable with those of Italian and French persons with HIV/AIDS. Accordingly, the expected number of cancer cases was computed based on national-based rates from population based cancer registries (see **Subheading 4.2.**). Furthermore, the standardization for sex, age, and area of origin was used to minimize the role of such factors in the computation of cancer risks. Although this epidemiologic study included [mt]11,000 individuals, for several cancer sites the statistical power was low, and it turned out to be sufficient for examining only the most common cancers in both groups (KS, NHL, lung and liver cancers).

Identification of cancer determinants and the quantification of the independent role of immunodepression is also problematic. Herein, individual data on immunosuppressive drugs were not available for most of the transplant recipients included in this study, and we were not able to assess the potential association between type of treatment and cancer risk. In addition, we lacked information on important risk factors other than immunosuppression (e.g., smoking, alcohol consumption, presence of chronic infections with viruses other than HIV).

4.2. Cancer Ascertainment and Computation of Expected Number of Cancer Cases

Potential incompleteness of follow-up also needs consideration. Most of study participants were not resident in areas covered by cancer registration; thus, the identification of new cases of cancer amongst the cohorts was based on the routine follow-up visits. This represents a limitation potentially affecting the internal validity of the study, because losses to follow-up could not be excluded. For example, it is possible that some patients, in view of symptoms/signs of a neoplastic disease, would refer to different physicians or centers and that some cancers, such as gynecological

cancers, may not be known by the infectious disease ward. With such a long study period, it is difficult to quantify the role of follow-up losses in cancer ascertainment, although for transplant recipients losses to follow-up are not likely to represent a strong limitation because these patients undergo strict medical surveillance to prevent organ rejection. Conversely, in HIV-infected patients, underestimation of new cancer cases is more likely to occur. A second important factor potentially influencing the study results and the comparison of cancer risks in these groups of immunosuppressed individuals is the choice of the referent population (i.e., the referent incidence rates from corresponding population-based cancer registries). For example, KS and NHL in Italy and France were extremely rare in young adults before the AIDS epidemic, whereas in recent years most of these cases were associated to HIV infection and, to a lesser extent, to organ transplant. Therefore, the reference incidence is influenced a lot by the incidence in these two groups and is not reflecting that much the incidence in the population excluding HIV and transplant patients *(14)*. In addition, because of that, the incidence in the reference data should have diminished considerably since the introduction of HAART. Moreover, for cancers in which the incidence varied considerably over the study period (such as lung cancer in women, or Hodgkin disease overall), the results might depend on the reference used.

Herein, we have use other data sources to avoid follow-up losses; however, incompleteness of follow-up among HIV-infected patients could still represent a drawback, and other tools (e.g., record linkage with cancer registries when possible) should be used to improve internal validity.

In conclusion, the cancer cohort consortium offers an efficient approach for studying the epidemiology of cancer given that several methodologic limitations are considered and dealt with. In southern Europe, this methodology used among immunosuppressed groups of population provided results that were in line with population based studies; however, further efforts are needed to improve the internal validity of this type of investigation.

Members of the Immunesuppression and Cancer Study Group: B. Longo, M Dorrucci, G. Rezza (Istituto Superiore di Sanità, Roma) for the Italian Seroconversion Study, Italy; M. P. Carrieri, C. Pradier for the DMI-2 Cohort Study Group, Nice University Hospital, France; A. Arbustini, B. Dal Bello, M. Grasso (Cardiochirurgia, Policlinico S. Matteo, IRCC, Pavia, Italy); F. Citterio, U. Pozzetto (Università Cattolica Sacro Cuore, Rome, Italy); A. Buda, P. Burra (Gastroenterologia, Padua University, Padua, Italy); U. Baccarrani (Chirurgia, Azienda Ospedale–Università, Udine, Italy); and G. Busnach, E. De Juli (Ospedale Niguarda Ca' Granda, Milan, Italy).

References

1. Doll R, Kinlen L (1970) Immunosurveillance and cancer: epidemiological evidence. *Br Med J* **4**, 420–422.
2. International Collaboration on HIV and Cancer (2000) The impact of highly active anti-retroviral therapy on the incidence of cancer in people infected with the human immunodeficiency virus. Collaborative reanalysis of individual data on 47,936 HIV-infected people from 23 cohort studies in 12 developed countries. *J Natl Cancer Inst* **92**, 1823–1830.
3. Biggar RJ, Chaturvedi AK, Goedert JJ, Engels AA. (2007) AIDS-related cancers and severity of immunosuppression in persons with AIDS. *J Natl Cancer Inst* **99**, 962–972.
4. Serraino D, Boschini A, Carrieri MP et al., (2000) Cancer risk among men with, or at risk of, HIV infection in southern Europe. *AIDS* **14**, 553–559.
5. Grulich AE, van Leuwen MT, O Falster M, M Vajdic C. (2007) Incidence of cancers in people with HIV/AIDS compared with immunosuppressed transplant recipients: a meta-analysis. *Lancet* **370**, 59–67.
6. Serraino D, Piselli P, Angeletti C, et al. (2005) Risk of Kaposi's sarcoma and of other cancers in Italian renal transplant patients. *B J Cancer* **92**, 572–575.
7. Pedotti P, Cardillo M, Rossini G, et al. (2003) Incidence of cancer after kidney transplant: results from the North Italy transplant program. *Transplantation* **76**, 1448–1451.
8. Herrero JL, Lorenzo M, Quiroga J, et al. (2005) De novo neoplasia after liver transplantation: an analysis of risk factors and influence on survival. *Liver Transpl* **11**, 89–97.
9. World Health Organization (1992) International Classification of Diseases and Related Problems, 10th rev. Geneva, Switzerland: World Health Organization.
10. Parkin DM, Whelan SL, Ferlay J, et al. (eds.) (1997) Cancer Incidence in Five Continents, Vol. VII. IARC Scientific Publications No. 143. Lyon, France: IARC Press.
11. Parkin DM, Whelan SL, Ferlay J, Teppo L, and Thomas DB (eds.) (2002) Cancer Incidence in Five Continents, Vol. VIII, IARC Scientific Publications No. 155. Lyon, France: IARC Press: Lyon.
12. Breslow NE, Day NE (1987) Statistical Methods in Cancer Research, Vol. 2. The design and analysis of cohort studies. IARC Scientific Publications No. 82. Lyon, France: :IARC Press.
13. Dal Maso L, Franceschi S, Polesel J, et al. (2003) Risk of cancer in persons with AIDS in Italy, 1985–98. **89**, 94–100.
14. Jones ME, Swerdlow AJ (1998) Bias in the standardized mortality ratio when using general population rates to estimate expected number of deaths. *Am J Epidemiol* **148**, 1012–1017.

Chapter 21

Do Viruses Cause Breast Cancer?

James S. Lawson

Summary

Because mouse mammary tumor virus (MMTV; the Bittner virus) is the proven cause of breast cancer in both field and experimental mice, similar viruses have long been suspects as a potential cause of human breast cancer. MMTV-like viral genetic material has been identified in human breast tumors, but there is no definitive evidence whether MMTV is causal and not merely an innocuous infection in humans. High-risk human papilloma viruses (HPVs), Epstein-Barr (EBV), and other viruses also have been identified in human breast tumors, but again there is no definitive evidenc e for a causal role. Any viral hypothesis as a cause of breast cancer must take into account the most striking epidemiologic feature of human breast cancer, the three- to sixfold differences in mortality and up to eightfold differences in incidence between some Asian and Western populations. These differences dramatically lessen to a two- to three-fold difference within one or two generations of migration of females from low to high risk of breast cancer countries. In this chapter, a plausible explanation for these phenomena is offered; that is, the hypothesis that oncogenic viruses such as MMTV and high-risk HPVs may initiate some breast cancers in most populations. Furthermore, dietary patterns are suggested to determine circulating sex hormone levels, which in turn promote the replication of the hormone-dependent viruses MMTV and HPV. In addition, diet and hormones promote growth of both normal and malignant cells. Finally, the hypothesis that migrants from low to high risk of breast cancer countries change their food consumption patterns is suggested, which leads to higher circulating hormone levels, which in turn promotes viral replication, which initiates breast oncogenesis, which is enhanced by sex and growth hormones.

Key words: Breast cancer, viruses, mouse mammary tumor virus, human papilloma virus, diet; food, hormones.

1. Introduction

The most striking epidemiologic feature of human breast cancer is the three- to sixfold differences in mortality and up to eightfold differences in incidence between some Asian and Western populations (*see* **Tables 21.1** and **21.2**). These differences dramatically

Mukesh Verma (ed.), *Methods in Molecular Biology, Cancer Epidemiology, vol. 471*
© 2009 Humana Press, a part of Springer Science + Business Media, Totowa, NJ
Book doi: 10.1007/978-1-59745-416-2

Table 21.1
Breast cancer mortality and food consumption. International comparisons

	Breast cancer mortality /100,000 females	Calories/day/capita	Total fats/capita
USA	19.0	3,774	157
UK	24.3	3,412	139
Australia	18.4	3,054	131
Japan	8.3	2,761	85
South Korea	4.4	3,058	77
China	5.5	2,951	90

Breast cancer mortality rates for year 2002/100,000 females age adjusted to world population. International Agency for Research on Cancer. CancerMondial. Globocan 2002 database (2005) http://www-dep.iarc.fr.
Apparent food consumption per capita as calories and total fats for year 2002 (Food and Agriculture Organization 2005).

Table 21.2
Breast cancer incidence per 100,000 (age adjusted), international comparisons for U.S., Chinese, and Japanese in home countries and for Chinese and Japanese born in home countries then migrated and living in the United States *(1)*

	U.S. women	Chinese in China	Chinese born in China migrated to USA	Chinese born in USA	Japanese in Japan	Japanese born in Japan migrated to USA	Japanese born in USA
Breast cancer incidence	159	19.1	47.2	58.9	19.7	41.1	74.5

Food consumption patterns directly influence the levels of cir culating hormones, which in turn influence breast cancer risk *(12)*. Also, it is well documented that migrants from low-income countries increase food consumption after migration to high-income countries—annual festival food gradually becomes daily food *(13)*.

lessen to a two- to threefold difference within one or two generations of migration of females from low to high risk of breast cancer countries *(1,2)*. These phenomena are shown in **Table 21.2** with respect to the incidence of breast cancer in the United States, China, and Japan compared with Chinese and Japanese

living in the United States. For Chinese and Japanese women born in Asia, the incidence doubles after migration to the United States. These data have been broadly consistent for >50 years *(3,4)*. There are exceptions to these trends as the premenopausal breast cancer incidence in Asian women has risen rapidly in recent decades *(5)*. These data strongly suggest that these differences in breast cancer incidence and mortality between populations are mainly due to external factors and are not genetically based.

What could be these external factors? The most obvious are the differing patterns of food consumption in various populations. Another possibility is differences in fertility patterns. Much less obvious is the possible role of oncogenic viruses. However, despite intensive research over many years, no conclusive evidence has emerged to identify any of these factors as having a major role in breast cancer etiology.

Conversely, there is conclusive evidence that a range of risk factors have a definite influence on breast cancer incidence. The most important of these is gender–the risk of breast cancer is approx. 150 times greater in females than males. The role of estrogens is crucial: basically no estrogens, no breast cancer *(6)*. The role of estrogens is complex, but it seems to be associated with the well known increased risk of breast cancer with early aged menarche, late age menopause, weight gain during and after menopause, and the use of postmenopausal hormone replacement therapy *(7)*. Alcohol consumption by women is also a small but accepted risk factor. Ionizing radiation, particularly when it affects young females, increases the risk of breast cancer *(8)*. Finally, full-term pregnancies at an early age, are associated with reduced risk of breast cancer.

There are strong and significant correlations between breast cancer mortality rates and calorie and fat consumption (**Table 21.1**). These correlations also have been consistent for >50 years. However, such correlations do not necessarily imply causation.

1.1. Working Hypothesis for a Hormone-Dependent Oncogenic Viral Etiology of Breast Cancer

There are >150,000 articles on some aspects of breast cancer that have been listed on Medline since 1966. The majority of these articles report some aspects of treatment, but many are concerned with etiology. It is of interest that these articles suggest that scientists seem to be as influenced by fashion as dress designers. During the 1960s and 1970s, viruses as possible causes of breast cancer was very fashionable; during the 1980s, fat and food consumption were favorites; and during the past 10 years, chasing gene pathways has become the dominant research activity.

Given the complexity of the risk factors for breast cancer, and this immense body of published data, it is helpful to develop hypotheses for the possible etiology of breast cancer. The creation of a range of hypotheses offers a logical way of considering this huge mass of fragmented evidence and in turn allows

research efforts to be more focused. Unfortunately, this approach also has lead to a proliferation of publications. In the past 12 months alone, there have been some 200 articles listed in Medline that are concerned with mostly different, causal hypotheses for breast cancer. The range of causal hypotheses is extraordinary and includes viruses, consumption of fats, ionizing radiation, exogenous and endogenous estrogens, insulin like growth factors, hormones and other influences on breast stem cells, alcohol, epigenetic and other gene factors, industrial residues such as organochlorines, influences during fetal life, and large breast size in slim women.

Because of the development of recent evidence, it is well worth revisiting and combining several of these hypotheses. This chapter outlines a food-influenced, hormone-dependent, viral etiology hypothesis for breast cancer. In brief, food consumption directly influences circulating hormone levels and hence tissue and overall growth in fetal, childhood, and adult life. Migrants from low- to high-income countries increase their food consumption in both volume and content and their children grow taller and heavier and their risk of breast cancer rapidly increases. Viruses with proven oncogenic potential, such as high-risk human papilloma viruses (HPVs) and mouse mammary tumor viruses (MMTVs), have been repeatedly identified in human breast tumors but rarely in normal breast tissues. Both HPVs and MMTVs have sex hormone-responsive gene sequences, and their replication is greatly increased in the presence of these hormones.

With respect to the influence of food on hormone metabolism, these hypotheses are in accord with those made >30 years ago by MacMahon et al. *(9)*. With respect to the influence of hormones on oncoviruses, these conclusions are the same as those of Varmus et al. *(10)*. As early as 1978, McGrath et al. *(11)* combined these hypotheses to form a cocarcinogenesis hypothesis for both hormones and mouse mammary tumor virus. A more detailed review of this evidence is presented below.

2. Food

The most obvious suspect external factor is food. As shown in **Table 21.1**, there are strong and significant correlations between total food and fat consumption and breast cancer mortality in both low and high risk for breast cancer countries. These correlations have been consistent for >50 years. This consistency suggests these trends are probably true despite the inevitable shortcomings in the methods and accuracy of the data collection.

However, there is substantial and consistent evidence that food consumption patterns, particularly of fats, do not influence

the risk of breast cancer in specific populations *(14)*. This conclusion has been reinforced by the recent intervention study of U.S. women in which a random sample of women consumed lower fat diets than a control sample of women *(15)*. There were no significant differences in the risk of invasive breast cancer despite the achievement of modest fat consumption in the intervention group. These observations in humans are in marked contrast to experimental observations in laboratory rodents in which there is consistent evidence that high-fat diets increase breast epithelial cell estrogen receptor expression and the incidence of mammary tumors *(16)*.

There is a possible explanation for this lack of correlation between food consumption patterns and breast cancer risk, namely, that food and fat consumption levels in study populations is above a threshold effect. Most studies have been of Western women among whom even the lowest food consumers exceed the highest fat and food consumers in selected Asian countries. Studies of food consumption and breast cancer risk among Indonesian women in Indonesia offers\ support for this hypothesis *(17)*. There is an eightfold difference in the odds ratios for breast cancer incidence among the highest compared with lowest fat consumption by presumed premenopausal Indonesian women. Similarly, Shanghai women who consume highest levels of fat have a near threefold increased risk of breast cancer compared with those women who consume lowest levels of fats *(18)*. The overall risk of breast cancer in both Indonesia and China is approximately one fifth that of Western women *(17,18)*.

2.1. Food and Hormones

It is likely that any influence of food on breast cancer risk is through its influence on hormone metabolism. There is consistent epidemiologic evidence that patterns of estrogen metabolism and serum estrogen levels are dependent on diet and that these are associated with breast cancer risk *(12,19)*. Urinary and serum estrogen and testosterone levels are >75% higher in Western compared with Asian women *(20)*. Serum sex hormone levels have been shown to be directly correlated with breast cancer risk in postmenopausal women, with a sixfold relative risk between highest and lowest hormone level quartiles of the study population *(21)*. Basically, no estrogens, no breast cancer, as shown by the low incidence of breast cancer among women who have been surgically castrated.

There is little evidence to suggest that food, estrogens, and other hormones are specifically carcinogenic, and animal models suggest that high levels of both saturated and unsaturated fats do not induce carcinogenesis unless a carcinogen is administered *(6,16)*.

Accordingly, it is likely that the risk factors for breast cancer that are associated with estrogens and other hormones—high

birth weight, early age menarche, late age menopause, and postmenopausal weight gain—are linked with patterns of food consumption. Some authors suggest that changes in food consumption patterns combined with changes in fertility patterns can alone account for the marked increase in breast cancer incidence among migrants to high risk for breast cancer countries *(8)*. This explanation is both likely and plausible for some of the increased risk after such migration, but it does not account for the whole of the increase. This increase requires additional explanation.

3. Viruses: Evidence for Possible Role of HPVs and MMTV in Human Breast Cancer

HPVs are specific to epithelial cells. Different HPV types are the proven cause of genital warts, genital condylomata, cervical cancer, and a range of head and neck cancers *(22,23)*.

3.1. HPVs and Breast Cancer

3.1.1. Animal Studies

Laboratory based studies aimed at determining the influence of hormones on HPVs is limited. However, virus-like particles containing the structural components of HPVs can be obtained. One region of the HPV genome contains genes designated E6 and E7, and they are involved in progression of normal cells to malignancy. Estrogen seems to play a role in HPV-related cervical cancer in mouse models *(24)*. Mouse kidney cell lines derived from transfection of HPV16 with EJ-ras or v-fos require the continued presence of hormones for proliferation *(25)*. Chronic estrogen exposure co-operates with HPV16 to cause carcinogenesis in the female reproductive tract of transgenic mice *(26)*. Transformation of primary rodent cells by HPV type 16 requires glucocorticoid hormones *(27)*.

3.1.2. Human Studies

Virtually all studies of the role of hormones in HPV-associated human oncogenesis have concerned cervical cancer. It is well established that steroid hormones promote the expression of high-risk HPVs in cervical cancer *(28,29)*. HPV infection alone is not sufficient to induce human cervical cancer. Steroid hormones such as estrogen and progesterone plus HPV infections are required for cervical epithelial cell transformation *(30)*. The biologic effects of glucocorticoids on HPV16-mediated human cell carcinogenesis are striking, with a 30-fold growth of immortalised cells after such exposure *(31)*. The expression of HPV16 transcripts increases eightfold by exposure of cervical cell lines to estrogens *(32)*. Estrogen and progesterone hormone receptors are highly expressed in HPV-infected cervical epithelial cells,

suggesting that sex steroid hormones are co-factors in HPV-related cervical neoplasia *(33)*.

There is a higher presence of HPV DNA in cervical smears of women with high compared to low levels of blood steroid hormone levels *(34)*.

The activation of HPV expression in cervical epithelial cells seems due to the specific interaction of glucocorticoid response elements in the HPV regulatory region of their genomes *(29,35,36)*.

3.1.2.1. HPV Viral Genetic Material is Present in Breast Tumour Tissues and Cells but Rarely in Normal Breast Tissues

The presence of HPV high risk types 16,18 and 33 in breast tumors has been identified in 10 of 12 studies including all published studies conducted since 1999 *(37–48)*. Normal breast tissue controls were available for 4 of these studies. In these 4 studies, there were 215 cases, HPV gene sequences were identified in 51 (23.7%). There were 89 controls and HPV sequences were identified in 1 (1.1%). In contrast to cervical cancer, HPVs are difficult to detect in breast cancer specimens. For example, in our own study, HPVs could not be detected in fresh frozen specimens using standard polymerase chain reaction (PCR), but could be detected with added amplification of DNA *(46)*. It seems likely that these difficulties may be due to low HPV DNA sequences in breast cancer.

3.1.2.2. Antibodies to HPVs in Sera and Tumours from Breast Cancer Patients and Controls

The presence of antibodies to HPV16 in the sera of patients with breast cancer is approx. 10%, which is no different to patients with non–HPV-related cancers *(49)*. The meaning of this observation is far from clear, because the presence of HPV16-associated breast and cervical cancer is well documented and antibodies should be present.

3.1.2.3. Malignant Transformation of Normal Breast Epithelial Cells by HPVs

Estrogens and other hormones promote the immortalisation of HPV-infected normal human breast epithelial cells *(50)*. For full malignant transformation of immortalised cells, accumulation of cellular changes by long-term passaging is necessary *(50)*.

The presence of HPV sequences in breast cancers have a typical malignant phenotype/morphology. HPVs are present in cancers occurring in human nipple milk ducts and these cancers have the typical histologic features of HPV-induced human cervical cancers *(45)*.

3.1.2.4. Methods of Transmission of HPVs

The means by which HPVs are transmitted to the breast are not known. Human papilloma virions are known to be released when the cornified envelope of cells desquamate; accordingly, HPVs can be transmitted by skin-to-skin contact as well as by sexual activity *(51)*. Transmission of HPVs is not confined to sexual activity. In a recent Finnish study, a range of high risk for cancer HPVs were identified in all family members, including infants *(52)*.

In addition, HPV sequences of the same type *(16)* have been identified in breast tumours that occur in women with HPV-associated cervical cancer *(41,44)*. Therefore, it is possible that

HPVs may be transmitted by hand from the female perineum to the breast, which could occur during sexual activities or even showering or bathing.

4. MMTV and Breast Cancer

MMTV is the well established etiologic agent of mammary tumors in field and experimental mice *(53)*. MMTV is transmitted both through mouse mother's milk (exogenous transmission) and through the mouse germline (endogenous transmission).

4.1. Animal-Based Studies

It has been rigorously established that MMTV is the cause of mouse mammary tumors (this term is used rather than breast cancer or tumors, because of the multiple mammary glands in mice and other rodents) *(54)*. Mouse mammary oncogenesis occurs because of the integration of MMTV into the host mouse genome. Although the MMTV genome does not contain any oncogenes, its integration promotes the expression of oncogenes. The mouse mammary tumors occur after a long latency period.

4.2. MMTV Is Markedly Influenced by Hormones

Estrogens induce mouse mammary tumors in the presence of MMTV but not in its absence *(55)*. Between 10 to 100 times more virus is produced by corticoid influenced MMTV-containing cells than by controls *(56,57)*. The DNA sequences involved in controlling glucocorticoid stimulation of MMTV transcription have been identified, and they are referred to as hormone-responsive elements *(58)*.

In the only study concerning the direct influence of food on MMTV-associated mouse mammary tumors, it was shown that restriction of energy intake significantly inhibited the development of mammary tumors *(59)*.

4.3. Human-Based Studies

The evidence is limited with respect to the influence of hormones on MMTV-like sequences in human-based studies. Sequences related to hormone-responsive elements in MMTV-like DNA have been identified in human breast cancers *(60)*. These hormone-responsive elements include glucocorticoid, estrogen receptor, progesterone, and prolactin response elements *(60)*.

The prevalence of MMTV-like sequences in human gestational breast cancer (cancer occurring during pregnancy or 12 months postpartum), is 62% compared with 30–38% of sporadic breast cancers, which suggests an influence of hormones on MMTV-like viruses *(60)*. The production of MMTV-like particles in the human breast cancer cell line T47D is stimulated by exposure to oestradiol and progesterone *(61)*.

4.4. MMTV-Like Viral Genetic Material Is Present in Breast Tumour Tissues and Cells but Rarely in Normal Breast Tissues

MMTV-like envelope gene sequences have been identified human breast cancer in 10 of 13 studies conducted since 1995 *(62–74)*. Most of these studies used primers identified by the Beatriz Pogo group as having low homology to human endogenous retroviral sequences *(62)*. Normal breast tissue controls (specimens from cosmetic surgery and benign nonproliferative breast conditions) were available for five of these studies. In these five studies, there were 612 breast cancer cases, MMTV-like gene sequences were identified in 222 (36.3%). There were 369 controls, MMTV-like sequences were identified in six (1.7%).

4.5. Specificity of MMTV

This virus has been identified in various organs of the mouse, but it is only oncogenic to mammary epithelial cells and to lymphocytes. The role of MMTV in humans seems to be confined to an oncogenic influence in breast epithelial cells and lymphomas. The reason for this specificity is not known.

In situ PCR-based studies have demonstrated the location of the MMTV-like env transcripts in breast cancer cells and not in surrounding lymphocytes or normal breast epithelial cells *(67)*.

The identification of MMTV-like sequences using PCR techniques is difficult, and some workers have been unable to identify any MMTV sequences in breast tumors *(66,68,70)*. However, using fluorescent tag PCR techniques, Zammarchi et al. *(74)* have been able to consistently identify MMTV-like gene sequences in 25% of human breast tumours.

Liu et al. *(75)* have amplified the complete proviral structure of MMTV-like virus, including the *env*-long terminal repeat, LTR gag, gag-pol and pol-env, sourced from each of two human breast carcinomas that were MMTV *env* positive. This 9.9-kb provirus was 95% homologous to MMTV. This provirus displayed typical features of a replication competent virus, plus the open reading frame for the superantigen and the glucocorticoid responsive element. Liu and colleagues argued that because these MMTV-like sequences are virtually undetectable in normal tissues and because they maintain open reading frames, they are most probably of exogenous origin. Recent work by Szabo et al. *(73)*, who were able to detect MMTV-like sequences in a variety of normal human tissues, has challenged this view, and they suggest that endogenous MMTV-like sequences also may exist *(73)*.

4.6. Infectivity of MMTV in Humans

The ability of MMTV to infect human cells is a prerequisite for a role in breast cancer, and although much indirect evidence has accumulated over the years, the recent direct demonstration of experimental infection of human cells by MMTV, together with the identification of MMTV–human genome junctions in two independently infected human cell lines of independent origin is of considerable interest *(76)*.

Although no evidence has yet been presented for experimental infection of human lymphocytes with MMTV, the presence of MMTV-like sequences in circulating lymphocytes in patients with breast cancer as well as in intestine lymphoid tissue has been reported *(77,78)*.

4.7. Detection of Antibodies to MMTV-Like Viruses and MMTV Antigens in Tumors from Breast Cancer Patients and Controls

After development of molecular techniques for the identification of MMTV and other viruses, interest in studies on MMTV-like antigens and the detection of MMTV-reactive antibodies lapsed, and many of the early studies are regarded as unsubstantiated by some observers (reviewed by Bar-Sinai et al. *[79]*). However, interest in the detection of MMTV antigens in human cancer as a marker was rekindled by the observation that MMTV p14 is translocated into the nucleoli in some human breast cancer specimens but not in normal breast tissue controls *(79)*.

In a recent investigation, no MMTV-specific antibodies were identified in women with breast cancer *(80)*. This observation challenges the concept of MMTV as a possible cause of human breast cancer. However, there is a possible explanation. If most new born human infants are exposed to MMTV via colostrum or breast milk, there may be a substantially reduced immune response as seems to be the case with new born mice *(81)*. It seems that virtually all newborn human infants are exposed to colostrum or milk soon after birth (NSW 2005) despite the apparent evidence to the contrary *(82)*. It may be that human newborn infants also have a reduced immune response to MMTV.

An MMTV-like superantigen also has been identified in human breast cancer *(83)*. This has parallels with the mouse model. In the mouse, expression of this superantigen is required for the stimulation of T lymphocytes, allowing amplification of virus infected cells and ultimately pathogenesis.

4.8. Transformation of Normal Breast Epithelial Cells by MMTV

MMTV, like all retroviruses, integrates randomly into the genome of infected cells. If MMTV integration of MMTV takes place in the vicinity of a cellular proto-oncogene, the expression of the proto-oncogene can be up-regulated. Analysis of a large number of murine mammary tumors has revealed that in almost all cases MMTV integration can be found either up or downstream of cellular proto-oncogenes of the *int* family. The expression of the cellular proto-oncogene is then up-regulated as a result of the influence of the strong MMTV transcriptional enhancer.

So far, it has not been possible to experimentally reproduce this transformation process in cell culture because a very large number of integration events are required before, by chance, the expression of a cellular proto-oncogene becomes deregulated. Furthermore, the overexpression of just one gene is not sufficient to lead to the fully transformed phenotype; rather, a number of genes need to be deregulated. However, for many years, it has

been suggested that MMTV itself may encode a protein that can also play a role in part of this transformation procedure and recent evidence from the laboratory of Susan Ross supports this notion *(84)*. This group was able to demonstrate that the Envelope protein of MMTV contains an immunoreceptor tyrosine-based activation motif, which, when expressed in normal human mammary cells growing in three-dimensional cultures, results in many changes associated with transformation *(84)*. However, these results have been achieved with MMTV, and they have not yet been demonstrated for human sourced MMTV-like virus.

Breast cancers carrying MMTV-like sequences have a typical malignant phenotype/morphology. It has been shown that 42.4% (n = 66) of human invasive breast tumor specimens may have some similar histologic characteristics to MMTV-induced mouse mammary tumor specimens *(85)*. However, there were no correlations between the presence and absence of MMTV sequences in the human breast tumors and similarity to mouse mammary tumors. It is unusual for the whole of the human breast tumor to show a similar histology to the mouse tumor, with similarities mostly being confined to only small areas of the human tumor.

4.9. Methods of Transmission of MMTV-Like Virus

In mice, it is known that MMTV is transmitted exogenously through mouse mother's milk and endogenously via the mouse germline. Although similar, there are several different strains of exogenous MMTV and early studies showed that their ability to cause mammary cancer in mice was different and also depended on the inbred strain of mouse that was infected with the virus.

MMTV-like particles have been observed in human milk from women with breast cancer and women with a family history of breast cancer *(86)*. These observations have not been confirmed or replicated and are regarded by many as controversial. However, using PCR, Ford *(87)* has identified MMTV-like *env* gene sequences in some human milk samples from normal Australian women.

Although there is seemingly consistent evidence that breast-fed babies are at no higher risk than nonbreast-fed babies of developing breast cancer, and at least among epidemiologists, there is a consensus that milk borne viruses are not associated with human breast cancer, spread of MMTV by human milk is a theoretical possibility *(82)*. This is because it has been realized that virtually all epidemiologic studies into breast feeding have been based on questionnaires, and mothers respond in varying ways depending on various understandings of the meaning of breast feeding. Current data, based on direct observation, suggest that it is possible that 90 to 100% of newborn babies are "exposed" to colostrum or breast milk despite failure of some mothers to continue breast feeding *(88)*. Accordingly transmission of MMTV by human breast milk to new born infants is possible.

It has been suggested that the worldwide distribution of MMTV-like virus gene sequences in breast cancer in human populations parallels the distribution of the MMTV-carrying common house mouse, *Mus domesticus(89)*. Continuous zoonotic (animal-to-human) transmissions of MMTV to humans are a possibility. Airborne particles of mouse materials are common household allergens and therefore are theoretical sources of transmission of MMTV from mice to humans. Transmission of MMTV by human ingestion of cereal and other food contaminated by mouse fecal material is also a speculative possibility.

5. Contrary Views

Mant et al. *(70)* have strongly argued that an MMTV-like virus is an unconvincing etiologic agent for human breast cancer. They challenge the proponents of MMTV-related human breast carcinogenesis to explain 1) how MMTV can infect *Homo sapiens* given that human cells lack the necessary transferring receptor for MMTV and that MMTV is not an endogenous human retrovirus? 2) Why immunosuppression does not increase the risk of breast cancer, as is the case with other human oncoviruses? 3) Why human breast feeding does not predispose daughters to breast cancer as is the case with mice? 4. Why pregnancy is protective against the risk of human breast cancer whereas the opposite is true for MMTV caused mouse mammary hyperplasia?

These seemingly valid points have been vigorously answered by Pogo and Holland *(90)*, the originators of the recent interest in MMTV-like viruses and human breast cancer. They and others argue 1) that MMTV has been shown to be able to infect human cells *(76)*; 2) four independent laboratories have identified MMTV-like *env* sequences in human breast tumors but rarely in normal control tissues and that negative findings by others may be due to use of different techniques; 3) that the MMTV-like provirus identified in human breast tumors is likely to be exogenous and not endogenous, primarily because the sequences have been found in breast tumors and rarely in normal breast tissues; and 4) in humans, early pregnancy is a protective factor against breast cancer, but gestational cancer (breast cancer during pregnancy and for 12 months postpartum) is extremely virulent suggesting that high levels of hormones may be involved. They finally argue that involvement of the immune system in MMTV infections is unique and differentiates this virus from other tumor viruses.

For this review, we suggest an additional speculative theory for the apparent lack of breast cancer in immunosuppressed patients

than in the general population: if MMTV infection in humans follows a similar route as in mice and involves an amplification of initially infected antigen-presenting cells via expression of a superantigen, then it would make sense that immunosuppression would have a major negative impact on virus amplification and thus on subsequent infection and transformation.

In addition, the striking international epidemiology of breast cancer, while almost certainly is associated with food consumption and related hormone patterns, could involve an unidentified epigenetic phenomena. Epigenetics refers to heritable changes in gene expression that occur without a change in gene sequences. There is limited evidence that such phenomena may be associated with an increased risk of breast cancer in some populations *(91)*.

6. Additional Viral Suspects

EBVs and Bovine leukemia viruses also may have role in breast cancer *(92)*. EBVs are probably not directly oncogenic but may act as cofactors. We hypothesize that EBVs may work in co-operation with MMTV. We base this hypothesis on the precedent of similar viral combinations for Kaposi sarcoma in which the retrovirus HIV and Kaposi herpes virus combine to cause oncogenesis. MMTV is a retrovirus and EBV is a herpes virus and a combined influence is theoretically possible.

7. General Discussion

This food, hormone, and viral co-carcinogenic hypothesis offers an explanation for virtually all of the known risk and other factors concerning breast cancer. However, the development of evidence that the scientific community would regard as conclusive, remains elusive.

There are criteria that have been developed to test the validity of the evidence that specific viruses cause specific cancers *(93)*. These criteria include the presence of viral genetic material in the cancer but rarely in normal tissues; the virus can transform or immortalize normal cells in culture; the presence of virus antibodies is more frequent in patients with the specific cancer than in control subjects; the presence of the virus, viral antibodies, or both preceded the development of the cancer; the virus leads to specific malignant phenotypes and characteristic histology; the oncogenicity of the virus in laboratory animals; and the positive

effect of an intervention (such as preventative vaccines) against the virus on the incidence of a specific cancer. This is a formidable set of criteria and in practice, none of the viruses that have been accepted as probable causes of cancer by the International Agency of Research in Cancer, including HPVs and cervical cancer, hepatitis viruses and liver cancer, Epstein-Barr viruses and lymphomas, meet all these criteria.

Conversely, both HPVs and MMTV with respect to their possible role in breast cancer, meet virtually all the above-mentioned criteria, with the important exceptions of evidence of a positive effect of an intervention plus consistent immunologic evidence. However, neither we nor other workers in this field, consider that the evidence is sufficiently persuasive to be conclusive of such causation. This is because the strength of any particular evidence is too low. This is in contrast to the evidence for the causal role of HPVs in cervical cancer. While that evidence is not as comprehensive as the list of criteria outlined above, the evidence is very strong because of the high HPV load and the global epidemiology.

Despite this lack of evidentiary strength, in our view HPVs and MMTV are major viral suspects as causes of human cancer. There is an urgent need to pursue additional evidence because of the possibility of primary prevention of breast cancer with vaccines for HPV and potential screening for MMTV.

References

1. Stanford J.L., Herrington L.J, Schwartz, S.M, Weiss, N.S. (1995) Breast cancer incidence in Asian migrants to the US and their descendants. *Epidemiology* **6**, 181–183.
2. Ziegler, R.G., Hoover R.N., Pike, M.C., Hidersheim A., Nomura, A.M.Y., et al. (1993) Migration patterns and breast cancer risk in Asian American women. *J Natl Cancer Instit* **85**, 1819–1827.
3. World Health Organization (2005) World Health Statistics Annual Report. World Health Organization, Geneva, Switzerland. 2005.
4. Althuis, M.D., Dozier J.M., Anderson, W.F., Devesa, S.S., Brinton, L.A. (2005) Global trends in breast cancer incidence and mortality 1973–1997. *Int J Epidemiol* **34**, 405–412.
5. Chia, K.S. Reilly, M. Tan, C.S. Lee, J. Pawitan, Y. Adami, H.O. Hall, P. Mow B. (2005) Profound changes in breast cancer incidence may reflect changes into a Westernized lifestyle: a comparative population-based study in Singapore and Sweden. *Int J Cancer* **113**, 302–306.
6. Yager, J.D., Davidson, N.E. (2006) Estrogen carcinogenesis in breast cancer. *N Engl J Med* **354**, 270–282.
7. Veronesi, U., Goldhirsch, A., Orecchia, R., Vlale, G, Boyle, P. (2005) Breast cancer. *Lancet* **365**, 1727–1741.
8. MacMahon, B. (2006) Epidemiology and the causes of breast cancer. *Int J Cancer* **118**, 2373–8.
9. MacMahon, B., Cole, P., Brown, J. (1972) Etiology of human breast cancer: a review. *J Natl Cancer Inst* **50**, 21–42.
10. Varmus, H.E., Ringold, G., Yamamoto, K.R. (1979) Regulation of mouse mammary tumor virus gene expression by glucocorticoid hormones. *Monogr Endocrinol* **12**, 253–278.
11. McGrath, C.M., Jones, R.F. (1978) Hormonal induction of mammary tumor viruses and its implications for carcinogenesis. *Cancer Res* **38**, 4112–4125.
12. MacMahon, B., Cole, P., Aoki, K., Lin, T.M., Morgan, R.W., Woo, N.C. (1974) Urine oestrogen profiles of Asian and

North American women. *Int J Cancer* **14**, 161–167.

13. Anderson, A.S., Bush, H., Lean, M., Bradby, H., Williams, R., Lea, E. (2005) Evolution of atherogenic diets in South Asian and Italian women after migration to a higher risk region. *J Hum Nutr Diet* **8**, 33–43.
14. Smith-Warner, S. A., Spiegelman, D., Adami, H. O., Beeson, W. L., van den Brandt, P. A., et al. (2001). Types of dietary fat and breast cancer: a pooled analysis of cohort studies. *Int J Cancer* **92**, 767–774.
15. Prentice, R. L., Caan, B., Chlebowski, R. T., Patterson, R., Kuller, L. H., et al (2006). Low-fat dietary pattern and risk of invasive breast cancer: the Women's Health Initiative Randomized Controlled Dietary Modification Trial. *JAMA* **295**, 629–642.
16. Hilakivi-Clarke L. (1997). Mechanisms by which high maternal fat intake during pregnancy increases breast cancer risk in female rodent offspring. *Breast Cancer Res Treat* **46**, 199–214.
17. Wakai, K., Dillon, D. S., Ohno, Y., Prihartono, J., Budiningsih, S., et al. (2000). Fat intake and breast cancer risk in an area where fat intake is low: a case control study in Indonesia. *Int. J. Epidemiol* **29**, 20–28.
18. Yuan, J.-M., Wang, Q.-S., Ross, R. K., Henderson, B. E., Yu, M.C. (1995). Diet and breast cancer in Shanghai and Tianjin, China. *Br J Cancer* **71**, 1353–1358.
19. Key, T. J., Allen, N. E., Spencer, E. A., Travis, R.C. (2002). The effect of diet on risk of cancer. *Lancet* **360**, 861–868.
20. Goldin, B. R., Adlercreutz, H., Gorbach, S. L., Woods, M. N., Dwyer, J. T., et al. (1986). The relationship between estrogen levels and diets of Caucasian American and Oriental immigrant women. *Am J Clin Nutr* **44**, 945–953.
21. Dorgan, J. F., Longcope, C., Stephenson, H. E., Falk, R. T., Miller, R., Franz, C., et al. (1997). Serum sex hormone levels are related to breast cancer risk in postmenopausal women. *Environ Health Perspect* **105(Suppl)**, 583–585.
22. International Agency for Research on Cancer (1995). Monographs on the Evaluation of Carcinogenic Risks to Humans. Human Papillomaviruses, Vol. 64. IARC, Lyon, France.
23. van Houten, V. M., Snijders, P.J. van den Brekel, M. W., Kummer, J. A., Meijer, C. J., et al. (2001). Biological evidence that human papillomaviruses are etiologically involved in a subgroup of head and neck squamous cell carcinomas. *Int J Cancer* **93**, 232–235.
24. Brake, T., Lambert, P.F. (2005). Estrogen contributes to the onset, persistence and malignant progression of cervical cancer in a human papilloma virus-transgenic mouse model. *Proc Natl Acad Sci USA* **102**, 2490–2495.
25. Crook, T., Storey, A., Almond, N., Osborn, K., Crawford, L. (1988). Human papillomavirus type 16 cooperates with activated ras and fos oncogenes in the hormone-dependent transformation of primary mouse cells. *Proc Natl Acad Sci USA* **85**, 8820–8824.
26. Arbeit, J M., Howley, P M., Hanahan, D. (1996). Chronic estrogen-induced cervical and vaginal squamous carcinogenesis in human papillomavirus type 16 transgenic mice. *Proc Natl Acad Sci USA*. **93**, 2930–2925.
27. Pater, M. M., Hughes, G. A., Hyslop, D. E., Nakshatri, H., Pater, A. (1988). Glucocorticoid-dependent oncogenic transformation by type 16 but not type 11 human papilloma virus DNA. *Nature*. **335**, 832–835.
28. Mittal, R., Pater, A., Pater, M.M. (1993). Multiple human papillomavirus type 16 glucocorticoid response elements functional for transformation, transient expression, and DNA-protein interactions. *J Virol* **67**, 5656–5659.
29. Khare, S., Pater, M.M.., Tang, S. C., Pater, A. (1997). Effect of glucocorticoid hormones on viral gene expression, growth, and dysplastic differentiation in HPV16-immortalized ectocervical cells. *Exp Cell Res* **232**, 353–360.
30. Webster, K., Taylor, A., Gaston, K. (2001). Oestrogen and progesterone increase the levels of apoptosis induced by the human papillomavirus type 16 E2 and E7 proteins. *J Gen Virol*. **82**, 201–213.
31. Khan, M. A., Canhoto, A. J., Housley, P. R., Creek, K. E., Pirisi, L. (1997). Glucocorticoids stimulate growth of human papillomavirus type 16 (HPV16)-immortalized human keratinocytes and support HPV16-mediated immortalization without affecting the levels of HPV16 E6/E7 mRNA. *Exp Cell Res* **236**, 304–310.
32. Mitrani-Rosenbaum, S., Tsvieli, R., Tur-Kaspa, R. (1989). Oestrogen stimulates differential transcription of human papillomavirus type 16 in SiHa cervical carcinoma cells. *J Gen Virol* **70**, 2227–2232.
33. Monsonego, J., Magdelenat, H., Catalan, F., Coscas, Y., Zerat, L., Sastre, X. (1991). Estrogen and progesterone receptors

in cervical human papillomavirus related lesions. *Int J Cancer* **48**, 533–539.

34. Kedzia, W., Gozdzicka-Jozefiak, A., Kwasniewska, A., Schmidt, M., Miturski, R., Spaczynski, M. (2000). Relationship between HPV infection of the cervix and blood serum levels of steroid hormones among pre- and postmenopausal women. *Eur J Gynaecol Oncol* **21**, 177–179.
35. Gloss, B., Bernard, H. U., Seedorf, K., Klock, G. (1987). The upstream regulatory region of the human papilloma virus-16 contains an E2 protein-independent enhancer which is specific for cervical carcinoma cells and regulated by glucocorticoid hormones. *EMBO J* **6**, 3735–3743.
36. Piccini, A., Storey, A, Romanos, M., Banks, L. (1997). Regulation of human papillomavirus type 16 DNA replication by E2, glucocorticoid hormone and epidermal growth factor. *J Gen Virol* **78**, 1963–1970.
37. Lonardo, D. A., Venuti, A., Marcante, M.L. (1992). Human papillomavirus in breast cancer. *Breast Cancer Res Treat* **21**, 95–100.
38. Bratthauer, G. L., Tavassoli, F.A. O'Leary, T.J. (1992). Etiology of breast carcinoma: no apparent role for papillomavirus types 6/11/16/18. *Pathol Res Pract* **188**, 384–386.
39. Wrede, D., Luqmani, Y. A., Coombes, R. C., Vousden, K.H. (1992). Absence of HPV 16 and 18 DNA in breast cancer. *Br J Cancer* **65**, 891–894.
40. Gopalkrishna, V., Singh, U. R., Sodhani, P., Sharma, J. K., Hedau, S. T., Mandal, A. K., Das B.C. (1996). Absence of human papillomavirus DNA in breast cancer as revealed by polymerase chain reaction. *Breast Cancer Res Treat* **39**, 197–202.
41. Hennig, E. M., Suo, Z., Thoresen, S., Holm, R., Kvinnsland, S., Nesland, J.M. (1999). Human papillomavirus 16 in breast cancer of women treated for high grade cervical intraepithelial neoplasia (CIN III). *Breast Cancer Res Treat* **53**, 121–135.
42. Yu, Y., Morimoto, T., Sasa, M., Okazaki, K., Harada, Y., Fujiwara, T., Irie, Y., Takahashi, E., Tanigami, A., Izumi, K. (1999). HPV33 DNA in premalignant and malignant breast lesions in Chinese and Japanese populations. *Anticancer Res* **19**, 5057–5061.
43. Damin, A.P.S., Karam, R., Zettler, C. G., Caleffi, M., Alexandre, C.O.P. (2004). Evidence for an association of human papillomavirus and breast carcinomas. *Breast Cancer Res Treat* **84**, 131–137.
44. Widschwendter, A., Brunhuber, T., Wiedemair, A., Mueller-Holzner, E., Marth, C. (2004). Detection of human papillomavirus DNA in breast cancer of patients with cervical cancer history. *J Clin Virol* **31**, 292–297.
45. De Villiers, E.-M., Sandstrom, R. E., zur Hausen, H., Buck, C.E. (2005). Presence of papillomatous sequences in condylomatous lesions of the mamillae and in invasive carcinoma of the breast. *Breast Cancer Res 7*, R1–R11.
46. Kan, C.-Y., Iacopetta, B. J., Lawson, J. S., Whitaker, N. J. (2005). Identification of human papillomavirus DNA gene sequences in human breast cancer. *Br J Cancer* **93**, 946–948.
47. Tsai, J. H., Tsai, C.H. Cheng, M. H., Lin, S. J., Xu, F.L. Yang, C.C. (2005). Association of viral factors with non-familial breast cancer in Taiwan by comparison with non-cancerous, fibroadenoma, and thyroid tumor tissues. *J Med Virol 75*, 276–81.
48. Kroupis, C., Markou, A., Vourlidis, N., Dionyssiou-Asteriou, A., Lianidou, E.S. (2006). Presence of high-risk human papillomavirus sequences in breast cancer tissues and association with histopathological characteristics. *Clin Biochem* **39**, 727–731.
49. Strickler, H. D., Schiffman, M. H., Shah, K.V,. Rabkin, C. S., Schiller, J. T., et al. (1998). A survey of human papillomavirus 16 antibodies in patients with epithelial cancers. Eur. J. Cancer Prev. 7, 305–313.
50. Dimri, G., Band, H., Band, V. (2005). Mammary epithelial cell transformation: insights from cell culture and mouse models. *Breast Cancer Res* 7, 171–179.
51. Bryan, J. T., Brown, D.R. (2001). Transmission of human papillomavirus type 11 infection by desquamated cornified cells. *Virology* **281**, 35–42.
52. Rintala, M.A.M., Grenman, S. E., Puranen, M. H., Isolauri, E., et al. (2005). Transmission of high risk human papillomavirus (HPV) between parent and infant: a prospective study of HPV in families in Finland. *J Clin Microbiol* **43**, 376–381.
53. Salmons, B., Miethke, T., Wintersperger, S., Muller, M., Brem, G., Gunzburg, W.H. (2000). Superantigen expression is driven by both mouse mammary tumor virus long terminal repeat-associated promoters in transgenic mice. *J Virol* **74**, 2900–2.
54. Moore, D. H., Long, C. A., Vaidya, A. A., Sheffield, J. B., Dion, A. S., Lasfargues, E.Y. (1979). Mammary tumor viruses. *Adv Cancer Res* **29**, 347–414.
55. Highman, B., Norvell, M. J., Shellenberger, T.E. (1977). Pathological changes in female

C3H mice continuously fed diets containing diethylstilboestrol or 17-estradiol. *J Environ Pathol Toxicol* **1**, 1–30.

56. Ringold, G. M., Lasgargues, E. Y., Bishop, J. M., Varmus, H.E. (1975). Production of MMTV by cultured cells in the absence and presence of hormones: assay by molecular hybridisation. *Virology* **65**, 135–147.
57. Svec, J, Links, J. (1977). MMTV production stimulated by hormones and polyamines in cells grown in semi-synthetic in vitro conditions. Int J Cancer **19**, 249–257.
58. Majors, J.E, Varmus, H.E. (1983). A small region of the mouse mammary tumor virus long terminal repeat confers glucocorticoid hormone regulation on a linked heterologous gene. *Proc Natl Acad Sci USA* **80**, 5866–5870.
59. Fernandes, G., Chandrasekar, B., Troyer, D. A., Venkatraman, J. T., Good, R.A. (1995). Dietary lipids and calorie restriction affect mammary tumor incidence and gene expression in mouse mammary tumor virus/v-Ha-ras transgenic mice. *Proc Natl Acad Sci USA* **92**, 6494–8.
60. Wang, Y., Melana, S. M., Baker, B., Bleiweiss, I., Fernandez-Cobo, M., Mandeli, J. F., Holland, J. F., Pogo, B.G.T. (2003). High prevalence of MMTV-like env gene sequences in gestational breast cancer. *Med Oncol* **20**, 233–236.
61. Faff, O., Murray, B. A., Erfle, V., Hehlmann, R. (1993). Large scale production and purification of human retrovirus-like particles related to the mouse mammary tumor virus. *FEMS Microbiol Lett* **109**, 289–296.
62. Wang, Y., Holland, J. F., Bleiweiss, I. J., Melana, S., Liu, X., et al. (1995). Detection of mammary tumor virus env gene-like sequences in human breast cancer. *Cancer Res* **55**, 5173–5151.
63. Etkind, P., Du, J., Khan, A., Pillitteri, J., Wiernik, P.H. (2000). Mouse mammary tumor virus-like *env* gene sequences in human breast tumors and a lymphoma of a breast cancer patient. *Clin Cancer Res* **6**, 1273–1278.
64. Melana, S. M., Holland, J. F., Pogo, B.G. (2001). Search for mouse mammary tumor virus-like env sequences in cancer and normal breast from the same individuals. *Clin Cancer Res* **7**, 283–4.
65. Melana, S. M., Picconi, M. A., Rossi, C., Mural, J., Alonio, L. V., Teyssie, A., et al. (2002). Detection of murine mammary tumor virus (MMTV) env gene like sequences in breast cancer from Argentine patients (Spanish). *Medicina 62*, 323–327.
66. Zangen, R., Harden, S., Cohen, D., Parrella, P., Sidransky, D. (2002). Mouse mammary tumor-like env gene as a molecular marker for breast cancer? *Int J Cancer* **102**, 304–307.
67. Ford, C. E., Tran, D., Deng, Y. M., Rawlinson, W. D., Lawson J.S. (2003). Mouse mammary tumour like virus prevalence in breast tumours of Australian and Vietnamese women. *Clin Cancer Res* **9**, 1118–1120.
68. Witt, A., Hartman, B., Marton, E., Zeillinger, R., Schrieber, M., Kubista E. (2003). The mouse mammary tumour-like env gene sequence is not detectable in the breast tissue of Austrian patients. *Oncol Rep* **10**, 1025–1029.
69. Ford, C. E., Faedo, M., Rawlinson, W.D. (2004). Mouse mammary tumor virus-like transcripts and DNA are found in affected cells of human breast cancer. *Clin Cancer Res* **10**, 7284–7289.
70. Mant, C., Cason, J. (2004). A human murine mammary tumour virus-like agent is an unconvincing aetiological agent for human breast cancer. *Rev Med Virol* **14**, 169–177.
71. Levine, P. H., Pogo, B.G.-T., Klouj, A., Coronel, S., Woodson, K., et al. (2004). Increasing evidence for a human breast carcinoma virus with geographic differences. *Cancer* **101**, 721–726.
72. Etkind, P. R., Stewart, A. F., Dorai, T., Purcell, D. J., Wiernik, P.H. (2004). Clonal isolation of different strains of mouse mammary tumor virus-like DNA sequences from both the breast tumors and non-Hodgkin's lymphomas of individual patients diagnosed with both malignancies. *Clin Cancer Res* **10**, 5656–5664.
73. Szabo, S., Haislip, A. M., Traina-Dorge, V., Costin, J. M., Crawford, B. E., et al. (2005). Human, rhesus macaque, and feline sequences highly similar to mouse mammary tumor virus sequences. *Microsc Res Tech* **68**, 209–221.
74. Zammarchi, F., Pistello, M., Piersigilli, A., Murr, R., Di Cristifanao, C., Naccarato, A. G., Bevilacqua, G. (2006). MMTV-like sequences in human breast cancer: a fluorescent PCR/laser microdissection approach. *J Pathol* **209**, 436–444.
75. Liu, B., Wang, Y., Melana, S. M., Pelisson, I., Najfield, V., Holland, J. F., Pogo, B.G.T. (2001). Identification of a proviral structure in human breast cancer. *Cancer Res* **61**, 1754–1759.
76. Indik, S., Guensburg, W. H., Salmons, B., Rouault, F. (2005). Mouse mammary

tumor virus infects human cells. *Cancer Res* **65**, 6651–6659.

77. Crepin, M., Lidereau, R., Chermann, J. C., Pouillar, T P., Magdamenat, H., Montagnier, L. (1984). Sequences related to mouse mammary tumor virus genome in tumor cells and lymphocytes from patients with breast cancer. *Biochem Biophys Res Commun* **118**, 324–331.
78. Lushnikova, A. A., Kryukova, I. N., Rotin, D. L., Lubchenko, L.N. (2004). Detection of the env MMTV-homologous sequences in mammary carcinoma patient intestine lymphoid tissue. *Doklady Biol Sci.* **399**, 423–426.
79. Bar-Sinai, A., Bassa, N., Fischette, M., Gottesman, M. M., Love, D. C., Hanover, J. A., Hochman, J. (2005). Mouse mammary tumor virus env-derived peptide associates with nucleolar targets in lymphoma, mammary carcinoma and human breast cancer. *Cancer Res* **65**, 7223–7230.
80. Goedert, J. J., Rabkin, C. S., Ross, S.R. (2006). Prevalence of serologic reactivity against four strains of mouse mammary tumour virus among US women with breast cancer. *Br J Cancer* **94**, 548–51.
81. Astori, M., Finke, D., Karapetian, O., Acha-Orbea, H. (1999). Development of T-B cell collaboration in neonatal mice. *Intern Immunol* **11**, 445–451.
82. Beral, V., Bull, D., Doll, D., Peto, R., Reeves, G. (2002). Breast cancer and breastfeeding: collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50 302 women with breast cancer and 96 973 women without the disease. *Lancet* **360**, 187–195.
83. Wang, Y., Jiang, J.-D., Xu, D., Li, Y., Qu, C., Holland, J. F., Pogo, B.G.-T. (2004). A mouse mammary tumor virus-like long terminal repeat superantigen in human breast cancer. *Cancer Res* **64**, 4105–4111.
84. Katz, E., Lareef, M. H., Rassa, J. C., Russo, J., Grande, S. M., et al. (2005). MMTV encodes an ITAM responsible for transformation of mammary epithelial cells in three-dimensional culture. *Journal of Experimental Medicine* **201**, 431–439.
85. Lawson, J. S., Tran, D. D., Carpenter, E., Ford, C. E., Rawlinson, W. D., Whitaker, N. J., Delprado, W. (2006). Presence of mouse mammary tumour-like virus gene sequences may be associated with specific human breast cancer morphology. *J Clin Pathol* **59**, 1287–1292.
86. Moore, D. H., Charney, J., Kramarsky, B., Lasfragues, E., Sarkar, N. H., et al. (1971). Search for a human breast cancer virus. *Nature* **229**, 611–615.
87. Ford, C.E. (2004). Viruses and breast cancer. Ph.D. thesis, University of New South Wales, Australia.
88. NSW (2005). Report on breast feeding in NSW 2001: NSW Department of Health, Sydney. www.cphn.biochem.usyd.edu.au.
89. Stewart, T.H.M., Sage, R. D., Stewart, A.F.R., Cameron, D.W. (2000). Breast cancer incidence highest in the range of one species of house mouse *Mus domesticus. Br J Cancer* **82**, 446–451.
90. Pogo, B.G.T., Holland, J.F. (2005). MMTV and human breast cancer. *Cancer Res* **65**, 1112.
91. Le Marchand, L., Haiman, C. A., Wilkens, L. R., Kolone, l L.N., Henderson, B.E. (2004). MTHFR polymorphisms, diet, HRT, and breast cancer risk: the multiethnic cohort study. *Cancer Epidemiol Biomarkers Prev* **13**, 2071–2077.
92. Lawson, J. S., Guenzburg, W., Whitaker, N.J. (2006). Viruses and breast cancer. Review. *Future Microbiol* **1**, 33–51.
93. Vonka, V. (2000). Causality in medicine: the case of tumours and viruses. *Philos Trans R Soc Lond* **355**, 1831–1841.

Chapter 22

Epidemiology of Human Papilloma Virus (HPV) in Cervical Mucosa

Subhash C. Chauhan, Meena Jaggi, Maria C. Bell, Mukesh Verma, and Deepak Kumar

Summary

In a worldwide scenario, human papillomavirus (HPV) infection is the second leading cause of cancer-related morbidity and mortality among women due to its very close association with cervical cancer. More than 100 different types of HPV genotypes have been characterized to date. Among these, approximately 24 HPV genotypes specifically infect the genital and oral mucosal system. The mucosal HPVs are most frequently sexually transmitted, and they are responsible for the most common sexually transmitted diseases throughout the world. In a majority of the cases, oncogenic/nononcogenic HPV infections spontaneously clear by themselves without any medical intervention. However, a persistent and long-term HPV infection usually leads to cervical cancer, which remains difficult to treat. In recent years, advance understanding of the structure of HPV and its pathogenesis has led to a variety of new treatments to combat HPV-related diseases, including a Food and Drug Administration-approved HPV vaccine that is very effective in young women. To effectively use this HPV vaccine worldwide, a clear understanding of HPV genotypes in different geographical populations is imperative. In this chapter, we have focused briefly on HPV genotypes and HPV prevalence in the women of different geographical populations.

Key words: Human papillomavirus, epigenetics, epidemiology, cervical cancer.

1. Cervical Cancer and Human Papillomavirus (HPV) Infection

According to statistical analyses released from the World Health Organization (WHO), cervical cancer is the second most common cancer in women worldwide *(1–3)*. Each year, approximately 493,000 new cases are diagnosed and 274,000 women die from cervical cancer worldwide *(4)*. The American Cancer Society estimates that in 2008, there will be 11.070 new diagnosed

Mukesh Verma (ed.), *Methods in Molecular Biology, Cancer Epidemiology, vol. 471*
© 2009 Humana Press, a part of Springer Science + Business Media, Totowa, NJ
Book doi: 10.1007/978-1-59745-416-2

cases of cervical carcinoma and 3,870 deaths associated with cervical cancer in the United States alone *(5)*. The incidence of cervical carcinoma is substantially higher among women of low socioeconomic status. Although Pap smear screening has decreased the incidence of cervical cancer globally, there are still pockets of the population, such as the American Indian women of the Northern Plains in United States, that have a significantly higher rate of cervical cancer *(6, 7)*. The known risk factors for cervical carcinoma include multiparity, smoking, immunosuppression, poor nutrition, and HPV infection *(8)*. However, the most important risk factors for cervical cancer are the persistence of an oncogenic HPV infection and a lack of timely screening *(9– 11)*.

2. Major HPV Types in Cervical Cancer

HPVs are members of a group of small nonenveloped DNA viruses. These viruses have a virion size of ~55 nm in diameter. More than 100 human HPV genotypes have been identified and sequenced thus far based on the sequence of their L1 genes, which differ from each other by at least 10% *(12)*. The presence of HPV DNA in cervical tissues has implicated HPV as a causative agent in genital condylomatas, in lower female genital tract intraepithelial neoplasias, such as cervical intraepithelial neoplasia (CIN), and in invasive cervical carcinomas *(13)*. It has been demonstrated that HPV DNA can be detected in approximately 99% of all invasive cervical cancers *(14)*. In addition, HPV DNA is almost always present in condylomatas and high-grade dysplasias, such as CIN III *(15)*. HPV types 6 and 11 are known to induce exophytic condylomatas affecting the anogenital mucosa and lower vagina *(16)*. A subset of HPV types (types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68) are regarded as oncogenic, or high-risk HPV viral types (**Table 22.1**). This subset represents the predominant HPV genotypes detected in high-grade intraepithelial lesions (CIN II and III) and in carcinomas of the lower female genital tract *(14,15,17)*.

A basic understanding of the HPV epidemiology is required to understand the role of various HPV types in the development of cervical cancer and to design effective vaccine strategies against the virus. The proportion of HPV infected women among different populations and across geographical regions may vary significantly. Additionally, different populations may harbor varying HPV genotypes in the genital tract *(14)*.

Table 22.1
Some common types of HPVs[a] associated to cervical cancer

High-risk types	Low-risk types
HPV16	HPV6
HPV18	HPV11
HPV26	HPV40
HPV31	HPV42
HPV33	HPV53
HPV35	HPV54
HPV39	HPV57
HPV45	HPV66
HPV51	HPV MM8[a]
HPV52	
HPV55	
HPV56	
HPV58	
HPV59	
HPV68	
HPV MM4	
HPV MM7	
HPV MM9	

HPVs can be classified into two main categories based on their malignant risk.

3. Molecular Mechanisms of HPV Infection in Causing Cervical Cancer

The HPVs are icosahedral capsids containing DNA viruses with a virion size of approximately 55 nm in diameter. This virus infects keratinocytes in the basal layer of cutaneous and mucosal epithelia, which replicates and assembles exclusively in the nuclei of the host cells (**Figs. 22.1A** and **B**). The HPV viruses are known to contain a small nonenveloped double-stranded DNA of approxi-

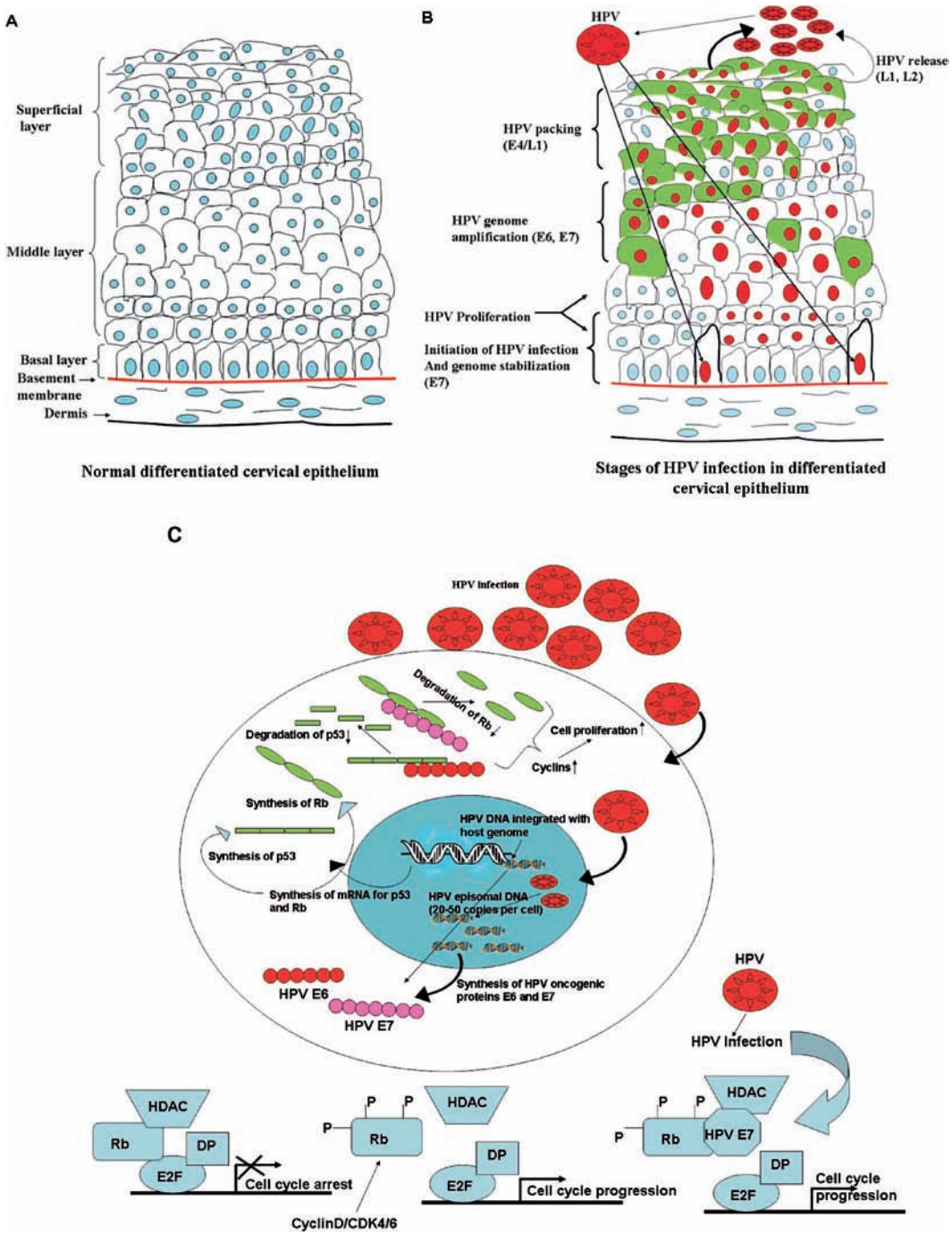


Fig. 22.1. Schematic presentation of cervical epithelium, orchestrated pattern of HPV infection and molecular mechanisms of oncogenic HPV proteins in cervical cancer progression. (**A**) The features of differentiated cervical epithelium are represented. The differentiated cervical epithelium is divided in three distinct layers, namely, basal layer or stratum basale, middle layer or stratum spongiosum, and superficial layer, which consists stratum granulosum and the uppermost stratum corneum. These three layers are required for completing HPV life cycle. (**B**) HPV infection initiates in the nuclei of basal layer cells and produces low copy number episomes, following expression of E7 HPV oncoprotein. Genome amplification and expression of E6 and E7 ocoproteins occur in middle layer. Following gene amplification, viral packing and release occur in upper layer, and E1–E4 proteins express by HPV. This orchestrated progression takes place only in a differentiated cervical epithelium. (**C**) Mode of HPV infection in host cell and molecular pathways modulate by HPV oncoproteins. After nuclear infection and generation of episomal copies, HPV genome concurrently integrates with host cell genome. The expression of E7 protein leads to perturb the cell cycle regulatory activities via interaction and subsequent proteosomal degradation of p53 and Rb proteins. Additionally, E7 protein also influences the activity of E2F promoter via interaction and phosphorylation of Rb and or HDAC proteins and titrating them away from E2F promoter. This leads to the activation of E2F promoter and cell cycle progression.

mately 8 kb. The HPV genome consists of three distinct parts, that is, early (6) and late (2) genes and a long control region (LCR) or noncoding region (NCR). The early genes encode for the expression of six nonstructural viral regulatory proteins (E1, E2, E4, E5, E6, and E7) in undifferentiated or intermediately differentiated keratinocytes. The late region of the genome encodes for the expression of two structural viral capsid proteins (L1 and L2) in keratinocytes undergoing terminal differentiation **(Fig. 22.1B)**. The roles of early regulatory genes are to overtake the host cell replication machinery, and the late genes encode for structural capsid proteins.

The two known oncoproteins produced by HPV are E6 and E7. The E6, a 150-amino acid HPV protein binds to p53 and targets its degradation via the ubiquitin pathway *(18)* for suppressing the proapoptotic activities of p53 and thereby promoting cell cycle progression **(Fig. 22.1C).** For degradation to occur, the E6 bound p53 (E6-p53) form a complex with ubiquitin protein (ligase) called E6AP *(19– 21)*. Formation of this complex targets p53 for the ubiquitination and its subsequent degradation via the 26S proteasome. This leads to a reduction in the half-life of p53 from several hours to <20 min in keratinocytes *(22– 24)*. Additionally, E6 also can indirectly inhibit p53 activity through its association with p300/CBP, which is a coactivator of p53 *(25– 27)*. The altered p53 expression/activity ultimately leads to cell proliferation, chromosomal duplication, and centrosomal abnormalities *(28– 32)*.

The expression of an approximately 100-amino acid HPV protein E7, however, primarily modulates the expression and activity of another cell cycle regulatory protein called retinoblastoma (Rb). The E7 protein binds and phosphorylates the Rb protein (pRb), causing its degradation **(Fig. 22.1C)** *(33)*. In a normal keratinocyte Rb-histone deacetylase (HDAC) protein complexes suppress the activity of E2F/DP-1 transcriptional complexes bound to the DNA in G1 phase of cell cycle. In S phase, Rb is phosphorylated by cyclin D/CDK4/6 kinases, which leads to the release of Rb-HDAC from E2F-DP1-DNA complex. The release of Rb-HDAC from the E2F-DP1-DNA complex leads in the activation of E2F-inducible genes (**Fig. 22.1C**) involved in DNA synthesis. In a high-risk HPV-infected keratinocyte, HPV E7 protein interacts with Rb and HDAC independently and titrates Rb and HDAC proteins away from the E2F-DP-1-DNA complex that lead to the constitutive activation of E2F-inducible genes (**Fig. 22.1C**). Additionally, E7 protein associates with cyclins A and E and cyclin-dependent kinase (cdk) inhibitors p21 and p27 *(34– 39)*. High-risk E7 proteins bind cyclinA-cdk2 complexes directly, and HPV18 E7 also binds cyclin E indirectly through p107 *(40)*. Moreover, two cdk inhibitors, p27 and p21, also have been shown to interact with E7 HPV protein, that may potentially block their action and thus may further enhance the activities of the cdks *(35– 37,39,41)*. By these mechanisms E7 protein induces cell cycle progression in HPV-infected cells

(42– 45). Epigenetic regulation of cervical cancer development in HPV-positive candidates also has been proposed in some studies *(46,47)*.

E6 and E7 are both highly expressed in cervical cancers. Therefore, they are considered to be excellent immunotherapy targets *(33,48)*. The causal relationship between HPV and cervical cancer has generated considerable interest in the development of vaccines for use in the treatment and prevention of cervical cancer. These HPV vaccines are type specific, meaning that the vaccines are designed to treat certain high-risk HPV types that infect the genital tract. Two distinct types of strategies have been used. The first of these strategies involves the development of prophylactic vaccines, which focus on the induction of effective humoral immune responses that may protect against HPV infection. The second of these strategies involves the development of therapeutic vaccines, which aim to stimulate cellular immune responses to eliminate virally infected cells. Based on extensive molecular studies on HPV16/18 a prophylactic vaccine has recently been developed which has received Food and Drub Administration approval for human use. However, before using the HPV vaccine(s) for a particular population, it is imperative to have relevant HPV epidemiology and genotyping data to provide an optimal outcome of vaccine(s) for that population.

4. Worldwide Epidemiology of HPV Infection

HPV infection is perhaps the most common sexually transmitted disease (STD) on the globe. According to a recent survey, it is estimated that the worldwide HPV prevalence among women is 10.5% (95% confidence interval [CI] = 9.9–11.0) at the present time *(14)*. The prevalence of HPV infection varies greatly in different populations and in different geographical regions. HPV prevalence is lowest in Spain (1.4%; 95% CI = 0.5–2.2) and highest in Nigeria (25.6%; 95% CI = 22.4–28.8) *(14)*. Epidemiologic and laboratory studies clearly suggest a very close association between HPV infection and cervical cancer. These data suggest that a basic understanding of the HPV epidemiology is required to predict expected prevalence of cervical cancer in different populations. Since >100 HPV types have been identified thus far and different populations may harbor varying HPV genotypes *(14)*, its also important to understand the prevalence of HPV genotypes. Thus, it is imperative to have relevant HPV genotyping data before using HPV vaccines for a particular population, to provide an optimal vaccine for providing the best possible care for that population. This chapter provides the baseline data that will be accessible to ensure that all population

can be appropriately included in vaccine trials in the future and to design effective population specific vaccine strategies. Based on available literature, we have determined the distribution of HPV infectivity in women of different continents and countries for the ease of understanding.

4.1. HPV Infection in Asia

On the basis of HPV infection prevalence, Asia is at number 3 in the world, with the highest (15.8%) prevalence in China followed by India (14%) and lowest in Vietnam Hanoi (1.6%). Herein, the reported HPV infection prevalence in China is based on two major HPV analyses reports, one report conducted in an urban area (Shenyang City; 685 women, aged 15–59 years) and the other report in the rural province of Shanxi, People's Republic of China (662 women, aged 15–59 years). Unlike most previous studies published on other populations, HPV prevalence among Chinese women was lower among women younger than 35 years (8.7%) than those older than 35 years (17.8%). However, like other populations, the most commonly found type was HPV16 (5.7% of all women and 38.8% of HPV-positive women), followed by HPV58, 52, 33, and 18 *(49,50)*. In Korea and Thailand, the HPV prevalence was reported as 8.5 and 6.0%, respectively. To investigate the prevalence of HPV infection in South Korea, a randomly selected sample of 863 sexually active women (age range, 20–74 years; median 44) and 103 self-reported virgins were collected and examined from the Busan region. The presence of DNA of 34 different HPV types in cervical exfoliated cells was tested by means of a polymerase chain reaction (PCR)-based assay. The overall prevalence of HPV DNA in sexually active Korean women was 10.4% (95% CI = 8.5–12.7). Surprisingly, the most commonly found HPV DNA type was HPV70 followed by HPV16 and HPV33 *(51)*. To investigate the prevalence and HPV infection and HPV genotypes in Thailand, a total of 1,741 women (age ≥15 years) were recruited from the Lampang and Songkla areas. HPV postitivity and genotypes were performed from the DNA isolated from exfoliated cervical cells by PCR. Of the total 1,741 women, 110 (6.3%) tested positive for the presence of HPV DNA. The overall age-standardized prevalence of HPV infection was higher (9.1%; 95% CI = 7.1–11.1) among the women (n = 1035) from Lampang compared with the women from Songkla (3.9%; 95% CI = 2.3–5.6; n = 706). HPV infection was more common in women <35 years old. The most common oncogenic HPV genotype was HPV16 followed by HPV52 and HPV72 *(52)*. To determine the HPV prevalence in India cervical cell samples were collected from 1,891women, aged 16–59 years residing in a rural area in the Dindigul District, Tamil Nadu, India. This is the single largest study conducted in India to determine the HPV prevalence. The overall 14.0% age-standardized HPV positivity was reported among Indian women. Among

all HPV-infected women, 21.9% of infections involved more than one HPV types. HPV16 (22.5% of women infected) was the most predominant high-risk HPV type followed by HPV56, HPV31, HPV33, and HPV18. Surprisingly, the HPV prevalence was constant across the age groups unlike most other populations studied in developed countries. However, the HPV positivity was inversely associated with education level (odds ratio [OR] among women with high school vs. no education = 0.6) *(53)*. Another major study on HPV infectivity in India that was aimed to investigated the differences in HPV prevalence in Hindu and Muslim (two major population types). For this study, 478 Indian Muslim and 534 Hindu women were recruited from the neighboring localities to avoid any regional differences and typed for HPV16 and 18. In this study, the overall HPV16/18 prevalence was 8.5%, and age-wise distribution of both subject-groups was similar. HPV16/18 infection in Muslim women was slightly higher (9.6%) compared with the Hindu women (7.5%). Among Muslim women HPV16/18 infections show a nonlinear trend trends that varied with age. In Hindu women, the prevalence is highest at age ≤24 years and linearly drops with increasing age *(54)*. Finally, the overall HPV prevalence in Asia was 8.3% and among all HPV-infected women HPV16 was the most common genotype. The presence of low-risk HPV types was lowest in Asia.

4.2. HPV Infectivity in Europe

The overall prevalence of HPV infection is relatively low in Europe, being highest in UK (10.9%) and lowest in Spain (1.3%). The HPV prevalence in the UK is based on a population-based study conducted in South Wales in relation to age, cytology, and social deprivation. For this study, DNA was purified from 1,911 liquid-based cytology samples (mean age 37.7 years; cytology 93.2% negative; social deprivation average score 17.9) and the presence of HPV DNA was determined by PCR-enzyme immunoassay (PCR-EIA). Of the 1911 women, 209 (10.9%) samples contained high-risk (HR) HPV infection. Among all HPV infected women, 36.4% women had multiple HPV infection. Like other population studies, the most common HPV genotype was HPV16 (19.6%) followed by HPV35 (9.5%), HPV66 (9.2%), HPV59 (8.5%), and HPV56 (7.6%). In this study, there was a strong association between HPV infection and cytologic abnormality. The infections of HR-HPVs were more frequent in younger women of age 30 years and under (68.9% of all HR-HPV infections Fisher's exact test, $P = 0.0001$) compared with the women of age 30 years and older. Overall, the data obtained from this study suggest a different HPV type distribution in South Wales compared with that reported for other populations *(55)*. Based on a random sample study of 1,013 women, aged 25–70 years, the overall HPV prevalence was 8.8% in Italy. The isolated DNA from cervical samples was tested for 36 HPV types by using

PCR with the general primers GP5+/GP6+. Among all HPV-infected women, HR genotypes were found in 7.1% of women with a multiple infection rate of only 1.1%. Of the all oncogenic HPV types, 32.6% of the women were infected with HPV16, and it was the most common HPV genotype found in Italian women. After age-adjustment, HPV prevalence was significantly higher in single women compared with married women (OR = 2.23; 95% CI = 1.28–3.89). Additionally, HPV prevalence was lower in parous women compared with the nulliparous women (OR = 0.49; 95% CI = 0.31–0.78). Overall, the Italian female population showed an intermediate HPV prevalence compared with other worldwide female populations *(56)*.

Two major studies were conducted to estimate the HPV prevalence in the general female Spanish population. Study was conducted in Barcelona (n = 973) by using PCR in 2003 and in 1992 another study in Basque Country (n = 1,178) by used slot-blot hybridization. Surprisingly, results of these two studies were very different. The Barcelona study used a random sample of female residents (with an the average age of 43 years) in metropolitan Barcelona, Spain, and we found the age-adjusted HPV prevalence was only 3.0% *(57)*. However, a 5.6 times higher (17%; 95% CI = 14–19) HPV prevalence was reported in the study conducted in Basque Country compared with the Barcelona study *(58)*. The main reason for low HPV prevalence in the Barcelona study may be participation of more women of age 30 and older (43 years was the average age of participants in this study). Because the results of these two studies show a large difference, it is difficult to draw a generalized conclusion on HPV prevalence in the general population in Spain. In a different European country, the Netherlands, a large study was conducted with 3,299 women to investigate HPV prevalence in the general population. Based on this study, 4.4% Dutch women showed HPV infectivity in the general population *(14)*.

The incidence and mortality rates of cervical cancer are higher in Germany compared with other Western European countries; however, the overall HPV prevalence was relatively lower according to a recent study *(59)*. In this study, type-specific HPV distribution was investigated in a large cohort of general women (n = 8,101) population in Germany. For this large study, women above age of 30 years were enrolled in two study centers in Hannover (northern Germany) and Tübingen (southern Germany). DNA samples isolated from the participants cervical samples were screened for HPV positivity by the hybrid capture 2 (HC2) test by using the high-risk probe. All samples that were positive by the HC2 test were genotyped using the prototype PGMY09/11 PCR line blot assay. HPV prevalence reported in this study was 6.4%, and prevalence of oncogenic types was 4.3%. Of the total PCR-positive women, 70.2% had a single infection, 28.1% had

multiple infections, and 1.7% remained uncharacterized. Among 32 different HPV types found, HPV16, 31, 52, 51, 18, and 45 were the most common oncogenic types in the German women. Although, in general, the overall HPV prevalence seems low in Europe, it is important to note that HPV16 (the most common oncogenic HPV genotype) infectivity was highest (21%) in this continent *(59)*.

4.3. HPV Infection Rate in North America

According to a recent Center for Disease Control report *(60)*, the prevalence of HPV infection is very high in the United States. This study was conducted by National Health and Nutrition Examination Survey. In this study, 1,921 U.S. female aged 14 to 59 years, provided a self-collected vaginal swab specimen in a mobile examination center. These swab samples were analyzed for the presence of HPV DNA by L1 consensus polymerase chain reaction followed by type-specific hybridization. Surprisingly, a very high overall HPV prevalence of 26.8% (95% CI = 23.3–30.9) was observed among the study participants. HPV prevalence was highest (44.8%; 95% CI = 36.3–55.3) among women aged 20 to 24 years, 24.5% (95% CI = 19.6–30.5) among females aged 14 to 19 years, 27.4% (95% CI = 21.9–34.2) among women aged 25 to 29 years, 27.5% (95% CI = 20.8–36.4%) among women aged 30 to 39 years, 25.2% (95% CI = 19.7–32.2) among women aged 40 to 49 years, and lowest (19.6%; 95% CI = 14.3–26.8) among women aged 50 to 59 years. The prevalence of HPV types 6 and 11 (low-risk types) and 16 and 18 (high-risk types) were detected in 3.4% of female participants; HPV6 was detected in 1.3% (95% CI = 0.8–2.3), HPV11 in 0.1% (95% CI = 0.03–0.3), HPV16 in 1.5% (95% CI = 0.9–2.6), and HPV18 in 0.8% (95% CI = 0.4–1.5) of female participants. These data clearly suggest that HPV infection is very common among females in the United States *(60)*. Another HPV prevalence study conducted in New Mexico by using cervical samples obtained from 3,863 women, aged 18–40 years old (mean 28 years) with no history of cervical abnormality reported an overall 39.2% HPV prevalence. Additionally, the prevalence of HR, low-risk, and uncharacterized HPV types was 26.7, 14.7, and 13.0%, respectively *(61)*. The high prevalence of HPV infection in U.S. women may be due to the lack of representation in the age group 60 years and older, which generally has a lower rate of HPV infectivity. Our laboratory recently conducted an HPV prevalence study in a general American Indian women population residing on Northern Plains by PCR by using the L1 consensus primer sets. To further examine the prevalence of specific genotypes the PCR products were hybridized with the Roche HPV Line Blot assay. In our study, of the total 287 women, 61 women (21.25%) tested positive for HPV infection. Among all HPV-positive women, 41 (67.2%) were infected with HR HPV

types. Of the HPV-infected women, 41% presented with multiple HPV genotypes. Additionally, of the women infected with oncogenic HPV types, 20 (48.7%) were infected with HPV16 and 18 and the remaining 21 (51.3%) were infected with other oncogenic types (i.e., HPV59, 39, 73) (unpublished findings). Overall these data suggest that the burden of HPV infection among U.S. females is greater than previous estimates.

Among Canadian women, the overall HPV prevalence is lower than in the United States. To determine HPV infectivity in Canadian women, an HPV analysis study was carried out in Ontario, among women of aged 15–49 years. Of the total 824 women, HPV infection was detected in 110 (13.3%; 95% CI = 11.0–15.6) women. Among all 824 women, 79 (9.6%; 95% CI = 7.6–11.6%) women were tested positive for carcinogenic HPV types by PCR and hybridization with type-specific probes. The prevalence of HPV was highest (24.0%; 95% CI = 16.5–31.5%) among women of age 20 to 24 years, and it was progressively decreased in older age groups, reaching 3.4% (95% CI = 0.1–6.7%) in women 45–49 years old similar to as observed in other populations *(62)*.

Among North American countries, cervical cancer is the most common cancer among Mexican women. To report the prevalence and determinants of HPV infection in Mexico, a population-based study was carried out between 1996 and 1999. In this study, 1,340 women in total with normal cytologic diagnoses were enrolled for HPV analysis. In Mexico the overall prevalence of HPV infection was 14.5% (95% CI = 14.4–14.6). In Mexican, women HPV infection showed a unique biphasic age distribution pattern and showed a highest (23%) infection rate in older women (age 65 years or older). Among Mexican women HPV infection was 16.7% in women of aged less than 25 years, with a rapid prevalence decline to 3.7% between 35 and 44 years. HPV infection increased again in the group of women aged between 45 and 54 years (12.3%), and it continued to increase to a maximum of 23% among women aged 65 years and older *(63)*. A similar HPV prevalence (16%) and age distribution pattern (biphasic or two peak age distribution pattern) was found in a large population-based study conducted in the female population of rural Costa Rica *(64)*. All these studies suggest a high prevalence of HPV infection in North America. Additionally, a different age-related distribution pattern of HPV infectivity was observed in this geographic region in some studies.

4.4. HPV Infection in South America

Based on three major studies conducted, in South America, the overall HPV prevalence is 14.3 being highest (16.3%) in Argentina and lowest in Chile (11.9%). In South American countries, HPV 16 showed the highest (15%) prevalence followed by HPV58 (7%). In a study aimed to determine HPV prevalence, the types

of HPV and age distribution, cervical specimens were collected from 1,038 normal sexually active women (age range, 15–69 years) from Santiago, Chile. The specimens were tested for the presence of HPV DNA by using a GP5+/6+ primer-mediated PCR and for cervical cytologic abnormalities by Papanicolaou smears. Of the total 1,038 women, 122 (11.7%) tested positive for HPV DNA. Among these HPV-infected women, 87 (71.3%) were positive with oncogenic HPV types, and 35 (28.7) were infected with nononcogenic HPV types. In total, 34 HPV types were found, including 13 oncogenic and 21 nononcogenic types *(65)*. In a different study conducted, aiming to identify type-specific prevalence of HPV infection in Columbia, 1,859 women in total from Bogota, Colombia, were tested for the presence of HPV DNA in their cervical specimen by using a general primer GP5+/GP6+ mediated PCR-EIA. The overall HPV DNA prevalence was 14.8% in Columbian women, and of the all HPV infections 9% of the women were infected by oncogenic HPV types. In total, 32 different HPV types were detected, with HPV16, 58, 56, 81, and 18 being the most common types. The HPV infection was highest (26.1%) prevalent among women younger than 20 years, lowest (2.3%) in women aged 45–54 years, and again peaked (13.2%) in women aged 55 years or more. Additionally, compared with women aged 35–44 years, women aged less than 20 years had a 10-fold increased risk of having multiple infections. Besides age, there was a positive association between the risk of HPV infection and number of regular sexual partners. In this population, multiplicity of sexual partners among young women, high educational level, and casual sexual partners seem to determine risk. *(66)*. In a similar study conducted in Argentina to evaluate HPV prevalence and determinants, cervical samples were collected from Concordia, from 1,786 women aged ≥15 years. These women were interviewed and underwent examination, including colposcopy. The presence of HPV DNA in cervical cells was tested with GP5+/6+ primer-mediated PCR-EIA. PCR was performed only on 987 specimens. Based on the presence of HPV DNA, a 16.3% prevalence of HPV infection was reported in Argentinian women. Among all HPV infections, HPV16 was the most (4.0%) common followed by HPV35 (2.6%). Additionally, almost half of the HPV-infected women showed multiple HPV infections *(67)*. Most of the data obtained from South American HPV prevalence studies suggest main risk factors for HPV infections are lifetime number of sex partners, sexual behavior, and vaginal discharge.

4.5. HPV Infectivity in Australia

Only a few studies have been published on HPV prevalence from the Australian continent. Based on two representative studies (one from Australia and one from New Zealand), the overall HPV prevalence in this continent is 18.9%. To determine HPV prevalence and

HPV genotype distribution in Australia, a small study was conducted in 282 women of Western Australia. DNA samples isolated from cervical samples from a cohort of 282 women were tested for the presence of HPV DNA by PCR followed by DNA sequencing to determine HPV genotypes. Of the 282 samples, 76 (27%) tested positive for HPV infection. Among all HPV-infected women, 39 (13.8%) were positive for oncogenic HPV types. In Australian women, HPV53 (15.8%) was the most common genotype followed by HPV16 (14.6%). These data suggest that HPV73 may be the main oncogenic HPV type in Australian women *(68)*. In another study, HPV prevalence was determined in Wellington women, to correlate presence of HPV infection with cervical cytology. In this study, 2,021 women in total attending family planning clinics in the Wellington region were enrolled to provide cervical samples. The presence of DNA in cervical scrapes was determined by dot-blot DNA hybridization. The mean age of participants was 26 years. In this study, 220 (10.9%) of women were infected with HPV. Among all HPV infections, HPV types 16, 18, or both were the most predominant infections, being detected in 71.5% women either alone or with other types. Multiple HPV infections were observed in 26.2% of the women *(69)*.

4.6. HPV Infection Rate in Africa

The incidence of cervical cancer in sub-Saharan Africa is among the highest in the world. The age-standardized rates are of 35.7 per 100,000 in Bamako, Mali, and 41.7 per 100,000 in Kyadondo, Uganda *(4)*. To investigate the risk factors for cervical cancer in this region, cervical samples from 932 sexually active women, aged 15 years or older, residing in city area of Ibadan, Nigeria, were obtained. The prevalence of HPV infection was determined in these women by PCR analysis and subsequent genotyping. In total, 32 different HPV genotypes were identified in Nigerian women. The overall HPV prevalence was 24.8% among women without cervical lesions. Oncogenic HPV types presented most frequently in this population were HPV 16, 31, 35, and 58. Among all HPV-infected women, 33.5% of women were infected with more than one HPV type. Unlike most populations studied so far, HPV prevalence was high not only among young women but also in middle-aged and old women. Additionally, single women (OR = 2.1; 95% CI = 1.1–3.9) and illiterate women (OR = 1.7; 95% CI = 1.1–2.5) showed higher rate of HPV infection. An overall high prevalence of HPV infection in all age groups was a distinctive feature of this population *(70)*. Presence of sustained HPV infection until up middle age and older may be a major risk factor for a higher incidence rate of cervical cancer in sub-Saharan African women. However, more HPV epidemiologic studies are warranted to draw a meaningful conclusion regarding this population.

4.7. HPV Infection in Antarctica

As such, there is no published literature available on HPV infection rate in women residing in Antarctica. However, some reports are available on HPV infection rate in Alaska Native women. In general, Alaska Native women have high rates of STDs and cervical cancer. The overall prevalence of HPV infections in Alaska Native women was 21%. These data were generated from a population-based study of 1,126 Alaska Native women in total. These women were seeking routine care and colposcopy examination in relation to cervical dysplasia. In addition to regular Pap test, cervical samples also were collected for HPV analysis by using a commercial dot hybridization test for seven HPV genotypes. Of the total 1,126 women, HPV DNA was detected in cervical specimens of 234 (21%) women. The prevalence of HPV DNA was declined with age and increased with sexual activity and cigarette smoking. However, it was associated with the use of oral contraceptives. Among all HPV infected women, 7.1% were infected with HR genotypes 16/18, and 14.4% of the women were infected with 31/33/35 HPV types *(71)*.

5. Conclusions

Epidemiologic and laboratory studies suggest that HPV infection is the main risk factor of cervical cancer worldwide. Examination of the worldwide prevalence of HPV demonstrates that the HPV infection rate is closely associated with 1) age of sexual debut, 2) number of sexual partners 3) younger age, and 4) personal hygiene. The distribution and prevalence of HPV infection varies in women of different geographical regions. Based on published reports, it is evident that HPV is maximally prevalent in the North American continent followed by the sub-Saharan African continent. HPV prevalence rate was lowest in European women; however, European women were maximally infected with HPV16 and 18 (accounting for 26% of all HPV infections). Studies also suggest that an HPV vaccine (that covers HPV6, 11, 16, and 18) will be maximally effective in European women, and it will have least coverage for HPV genotypes presented in North American women.

Acknowledgments

We acknowledge Cathy Christopherson for editorial assistance. This research was supported by a Sanford Research/USD grant and a Sanford School of Medicine research grant. D.K. is supported by MBRS SCORE grant from National Institute of General Medical Sciences.

References

1. Pisani, P., Bray, F., and Parkin, D. M. (2002). Estimates of the world-wide prevalence of cancer for 25 sites in the adult population. *Int J Cancer* **97**, 72–81.
2. Pisani, P., Parkin, D. M., Bray, F., and Ferlay, J. (1999). Estimates of the world-wide mortality from 25 cancers in 1990. *Int J Cancer* **83**, 18– 29.
3. Iihara, K., Iihara K, Shiozaki H, Tahara H, Kobayashi K, Inoue M, Tamura S, et al. (1993). Prognostic significance of transforming growth factor alpha in esophageal carcinoma. *Cancer* **71**, 2902–2909.
4. Parkin, D. M., Bray, F., Ferlay, J., and Pisani, P. (2005). Global cancer statistics, 2002. *CA Cancer J Clin* **55**, 74–108.
5. Cancer Facts and Figures. 2008.
6. Cobb, N., and Paisano, R. (1997). Cancer mortality among American Indians and Alaska Natives in the United States: regional differences in Indian Health, 1989–1993. Rockville, MD: Indian Health Service.
7. Leman, R. F., Espey, D., and Cobb, N. (2005). Invasive cervical cancer among American Indian women in the Northern Plains, 1994–1998: incidence, mortality, and missed opportunities. *Public Health Rep* **120**, 283–7.
8. Hoskins, W. J., Perez, C. A., and Young, R. C. (eds.) (2000). Principles and Practice of Gynecologic Oncology. Philadelphia, PA: Lippincott Williams and Wilkins.
9. Damasus-Awatai, G., and Freeman-Wang, T. (2003). Human papilloma virus and cervical screening. *Curr Opin Obstet Gynecol* **15**, 473–7.
10. Reid, J. (2001). Women's knowledge of Pap smears, risk factors for cervical cancer, and cervical cancer. *J Obstet Gynecol Neonatal Nurs* **30**, 299–305.
11. Castle, P. E., Shields, T., Kirnbauer, R., Manos, M. M., Burk, R. D., Glass, A. G., Scott, D. R., Sherman, M. E., and Schiffman, M. (2002). Sexual behavior, human papillomavirus type 16 (HPV 16) infection, and HPV 16 seropositivity. *Sex Transm Dis* **29**, 182–7.
12. de Villiers, E. M., Fauquet, C., Broker, T. R., Bernard, H. U., and zur Hausen, H. (2004). Classification of papillomaviruses. *Virology* **324**, 17–27.
13. Castle, P. E., Hillier, S. L., Rabe, L. K., Hildesheim, A., Herrero, R., Bratti, M. C., Sherman, M. E., Burk, R. D., Rodriguez, A. C., Alfaro, M., Hutchinson, M. L., Morales, J., and Schiffman, M. (2001). An association of cervical inflammation with high-grade cervical neoplasia in women infected with oncogenic human papillomavirus (HPV). *Cancer Epidemiol Biomarkers Prev* **10**, 1021–7.
14. Clifford, G. M., Gallus, S., Herrero, R., Munoz, N., Snijders, P. J., Vaccarella, S., Anh, P. T., Ferreccio, C., Hieu, N. T., Matos, E., Molano, M., Rajkumar, R., Ronco, G., de Sanjose, S., Shin, H. R., Sukvirach, S., Thomas, J. O., Tunsakul, S., Meijer, C. J., and Franceschi, S. (2005). Worldwide distribution of human papillomavirus types in cytologically normal women in the International Agency for Research on Cancer HPV prevalence surveys: a pooled analysis. *Lancet* **366**, 991–8.
15. Richardson, H., Franco, E., Pintos, J., Bergeron, J., Arella, M., and Telier, P. (2000). Determinants of low-risk and high-risk cervical human papillomavirus infectionsn in Montreal University students. *Sex Trans Dis* **27**, 79–86.
16. Franco, E. (ed.) (1996). Epidemiology of anogenital warts and cancer. Obstetrics and Gynecology clinics of North America: Human Papillomavirus I. Philadelphia, PA: W. B. Saunders Company.
17. ZurHausen, H., and Villiers, E. (1994). Human papillomavirus. *Annu Rev Microbiol* **48**, 427–447, 1994.
18. Werness, B., Levine, A., and Howley, P. (1990). Association of human papillomavirus types 16 and 18 E6 proteins with p53. *Science* **248**, 76–79.
19. Huibregtse, J. M., Scheffner, M., and Howley, P. M. (1991). A cellular protein mediates association of p53 with the E6 oncoprotein of human papillomavirus types 16 or 18. *EMBO J* **10**, 4129–35.
20. Howley, P. M., Munger, K., Romanczuk, H., Scheffner, M., and Huibregtse, J. M. (1991). Cellular targets of the oncoproteins encoded by the cancer associated human papillomaviruses. *Princess Takamatsu Symp* **22**, 239–48.
21. Howley, P. M., Scheffner, M., Huibregtse, J., and Munger, K. (1991). Oncoproteins encoded by the cancer-associated human papillomaviruses target the products of the retinoblastoma and p53 tumor suppressor genes. *Cold Spring Harb Symp Quant Biol* **56**, 149–55.
22. Huibregtse, J. M., Scheffner, M., and Howley, P. M. (1993). Cloning and expression of the cDNA for E6-AP, a protein that mediates

the interaction of the human papillomavirus E6 oncoprotein with p53. *Mol Cell Biol* **13**, 775–84.

23. Hubbert, N. L., Sedman, S. A., and Schiller, J. T. (1992). Human papillomavirus type 16 E6 increases the degradation rate of p53 in human keratinocytes. *J Virol* **66**, 6237–41.
24. Sedman, S. A., Hubbert, N. L., Vass, W. C., Lowy, D. R., and Schiller, J. T. (1992). Mutant p53 can substitute for human papillomavirus type 16 E6 in immortalization of human keratinocytes but does not have E6-associated trans-activation or transforming activity. *J Virol* **66**, 4201–8.
25. Lechner, M. S., and Laimins, L. A. (1994). Inhibition of p53 DNA binding by human papillomavirus E6 proteins. *J Virol* **68**, 4262–73.
26. O'Connor, M. J., Zimmermann, H., Nielsen, S., Bernard, H. U., and Kouzarides, T. (1999). Characterization of an E1A-CBP interaction defines a novel transcriptional adapter motif (TRAM) in CBP/p300. *J Virol* **73**, 3574–81.
27. Zimmermann, H., Degenkolbe, R., Bernard, H. U., and O'Connor, M. J. (1999). The human papillomavirus type 16 E6 oncoprotein can down-regulate p53 activity by targeting the transcriptional coactivator CBP/p300. *J Virol* **73**, 6209–19.
28. Demers, G. W., Foster, S. A., Halbert, C. L., and Galloway, D. A. (1994). Growth arrest by induction of p53 in DNA damaged keratinocytes is bypassed by human papillomavirus 16 E7. *Proc Natl Acad Sci U S A* **91**, 4382–6.
29. Foster, S. A., Demers, G. W., Etscheid, B. G., and Galloway, D. A. (1994). The ability of human papillomavirus E6 proteins to target p53 for degradation in vivo correlates with their ability to abrogate actinomycin D-induced growth arrest. *J Virol* **68**, 5698–705.
30. Galloway, D. A., Demers, G. W., Foster, S. A., Halbert, C. L., and Russell, K. (1994). Cell cycle checkpoint control is bypassed by human papillomavirus oncogenes. *Cold Spring Harb Symp Quant Biol* **59**, 297–306.
31. Kessis, T. D., Slebos, R. J., Nelson, W. G., Kastan, M. B., Plunkett, B. S., Han, S. M., Lorincz, A. T., Hedrick, L., and Cho, K. R. (1993). Human papillomavirus 16 E6 expression disrupts the p53-mediated cellular response to DNA damage. *Proc Natl Acad Sci U S A* **90**, 3988–92.
32. Thompson, D. A., Belinsky, G., Chang, T. H., Jones, D. L., Schlegel, R., and Munger, K. (1997). The human papillomavirus-16 E6 oncoprotein decreases the vigilance of mitotic checkpoints. *Oncogene* **15**, 3025–35.
33. Dyson, N., Howley, P., and Munger, K. (1989). The human papillomavirus-16 oncoprotein is able to bind the retinoblastoma gene product. *Science* **243**, 934.
34. Davies, R., Hicks, R., Crook, T., Morris, J., and Vousden, K. (1993). Human papillomavirus type 16 E7 associates with a histone H1 kinase and with p107 through sequences necessary for transformation. *J Virol* **67**, 2521–8.
35. Funk, J. O., Waga, S., Harry, J. B., Espling, E., Stillman, B., and Galloway, D. A. (1997). Inhibition of CDK activity and PCNA-dependent DNA replication by p21 is blocked by interaction with the HPV-16 E7 oncoprotein. *Genes Dev* **11**, 2090–100.
36. Jones, D. L., Alani, R. M., and Munger, K. (1997). The human papillomavirus E7 oncoprotein can uncouple cellular differentiation and proliferation in human keratinocytes by abrogating p21Cip1-mediated inhibition of cdk2. *Genes Dev 11*, 2101–11.
37. Jones, D. L., and Munger, K. (1996). Interactions of the human papillomavirus E7 protein with cell cycle regulators. *Semin Cancer Biol* **7**, 327–37.
38. Tommasino, M., Adamczewski, J. P., Carlotti, F., Barth, C. F., Manetti, R., Contorni, M., Cavalieri, F., Hunt, T., and Crawford, L. (1993). HPV16 E7 protein associates with the protein kinase p33CDK2 and cyclin A. *Oncogene* **8**, 195–202.
39. Zerfass-Thome, K., Zwerschke, W., Mannhardt, B., Tindle, R., Botz, J. W., and Jansen-Durr, P. (1996). Inactivation of the cdk inhibitor p27KIP1 by the human papillomavirus type 16 E7 oncoprotein. *Oncogene* **13**, 2323–30.
40. McIntyre, M. C., Ruesch, M. N., and Laimins, L. A. (1996). Human papillomavirus E7 oncoproteins bind a single form of cyclin E in a complex with cdk2 and p107. *Virology* **215**, 73–82.
41. Jones, D. L., and Munger, K. (1997). Analysis of the p53-mediated G1 growth arrest pathway in cells expressing the human papillomavirus type 16 E7 oncoprotein. *J Virol* **71**, 2905–12.
42. Longworth, M. S., and Laimins, L. A. (2004). Pathogenesis of human papillomaviruses in differentiating epithelia. *Microbiol Mol Biol Rev* **68**, 362–72.
43. Longworth, M. S., and Laimins, L. A. (2004). The binding of histone deacetylases and the integrity of zinc finger-like motifs of the E7 protein are essential for the life cycle

of human papillomavirus type 31. *J Virol* **78**, 3533–41.

44. Oh, S. T., Longworth, M. S., and Laimins, L. A. (2004). Roles of the E6 and E7 proteins in the life cycle of low-risk human papillomavirus type 11. *J Virol* **78**, 2620–6.
45. Khor, T. O., Keum, Y. S., Lin, W., Kim, J. H., Hu, R., Shen, G., Xu, C., Gopalakrishnan, A., Reddy, B., Zheng, X., Conney, A. H., and Kong, A. N. (2006). Combined inhibitory effects of curcumin and phenethyl isothiocyanate on the growth of human PC-3 prostate xenografts in immunodeficient mice. *Cancer Res* **66**, 613–21.
46. Turan, T., Kalantari, M., Cuschieri, K., Cubie, H. A., Skomedal, H., and Bernard, H. U. (2007). High-throughput detection of human papillomavirus-18 L1 gene methylation, a candidate biomarker for the progression of cervical neoplasia. *Virology* **361**, 185-93,
47. de la Cruz-Hernandez, E., Perez-Cardenas, E., Contreras-Paredes, A., Cantu, D., Mohar, A., Lizano, M., and Duenas-Gonzalez, A. (2007). The effects of DNA methylation and histone deacetylase inhibitors on human papillomavirus early gene expression in cervical cancer, an in vitro and clinical study. *Virol J* **4**, 18.
48. Crish, J. F., Bone, F., Balasubramanian, S., Zaim, T. M., Wagner, T., Yun, J., Rorke, E. A., and Eckert, R. L. (2000). Suprabasal expression of the human papillomavirus type 16 oncoproteins in mouse epidermis alters expression of cell cycle regulatory proteins. *Carcinogenesis* **21**, 1031–7.
49. Li, L. K., Dai, M., Clifford, G. M., Yao, W. Q., Arslan, A., Li, N., Shi, J. F., Snijders, P. J., Meijer, C. J., Qiao, Y. L., and Franceschi, S. (2006). Human papillomavirus infection in Shenyang City, People's Republic of China: A population-based study. *Br J Cancer* **95**, 1593–7.
50. Dai, M., Bao, Y. P., Li, N., Clifford, G. M., Vaccarella, S., Snijders, P. J., Huang, R. D., Sun, L. X., Meijer, C. J., Qiao, Y. L., and Franceschi, S. (2006). Human papillomavirus infection in Shanxi Province, People's Republic of China: a population-based study. *Br J Cancer* **95**, 96–101.
51. Shin, H. R., Lee, D. H., Herrero, R., Smith, J. S., Vaccarella, S., Hong, S. H., Jung, K. Y., Kim, H. H., Park, U. D., Cha, H. S., Park, S., Touze, A., Munoz, N., Snijders, P. J., Meijer, C. J., Coursaget, P., and Franceschi, S. (2003). Prevalence of human papillomavirus infection in women in Busan, South Korea. *Int J Cancer* **103**, 413–21.
52. Sukvirach, S., Smith, J. S., Tunsakul, S., Munoz, N., Kesararat, V., Opasatian, O., Chichareon, S., Kaenploy, V., Ashley, R., Meijer, C. J., Snijders, P. J., Coursaget, P., Franceschi, S., and Herrero, R. (2003). Population-based human papillomavirus prevalence in Lampang and Songkla, Thailand. *J Infect Dis* **187**, 1246–56.
53. Franceschi, S., Rajkumar, R., Snijders, P. J., Arslan, A., Mahe, C., Plummer, M., Sankaranarayanan, R., Cherian, J., Meijer, C. J., and Weiderpass, E. (2005). Papillomavirus infection in rural women in southern India. *Br J Cancer* **92**, 601–6.
54. Duttagupta, C., Sengupta, S., Roy, M., Sengupta, D., Bhattacharya, P., Laikangbam, P., Roy, S., Ghosh, S., and Das, R. (2004). Are Muslim women less susceptible to oncogenic human papillomavirus infection? A study from rural eastern India. *Int J Gynecol Cancer* **14**, 293–303.
55. Hibbitts, S., Rieck, G. C., Hart, K., Powell, N. G., Beukenholdt, R., Dallimore, N., McRea, J., Hauke, A., Tristram, A., and Fiander, A. N. (2006). Human papillomavirus infection: an anonymous prevalence study in South Wales, UK. *Br J Cancer* **95**, 226–32.
56. Ronco, G., Ghisetti, V., Segnan, N., Snijders, P. J., Gillio-Tos, A., Meijer, C. J., Merletti, F., and Franceschi, S. (2005). Prevalence of human papillomavirus infection in women in Turin, Italy. *Eur J Cancer 41*, 297–305.
57. de Sanjose, S., Almirall, R., Lloveras, B., Font, R., Diaz, M., Munoz, N., Catala, I., Meijer, C. J., Snijders, P. J., Herrero, R., and Bosch, F. X. (2003). Cervical human papillomavirus infection in the female population in Barcelona, Spain. *Sex Transm Dis* **30**, 788–93.
58. Mugica-Van Herckenrode, C., Malcolm, A. D., and Coleman, D. V. (1992). Prevalence of human papillomavirus (HPV) infection in Basque Country women using slot-blot hybridization: a survey of women at low risk of developing cervical cancer. *Int J Cancer 51*, 581–6.
59. Klug, S. J., Hukelmann, M., Hollwitz, B., Duzenli, N., Schopp, B., Petry, K. U., and Iftner, T. (2007). Prevalence of human papillomavirus types in women screened by cytology in Germany. *J Med Virol* **79**, 616–25.
60. Dunne, E. F., Unger, E. R., Sternberg, M., McQuillan, G., Swan, D. C., Patel, S. S., and Markowitz, L. E. (2007). Prevalence of HPV infection among females in the United States. *JAMA* **297**, 813–9.
61. Peyton, C. L., Gravitt, P. E., Hunt, W. C., Hundley, R. S., Zhao, M., Apple, R. J., and Wheeler, C. M. (2001). Determinants

of genital human papillomavirus detection in a US population. *J Infect Dis* **183**, 1554–64.

62. Sellors, J. W., Karwalajtys, T. L., Kaczorowski, J., Mahony, J. B., Lytwyn, A., Chong, S., Sparrow, J., and Lorincz, A. (2003). Incidence, clearance and predictors of human papillomavirus infection in women. *CMAJ* **168**, 421–5.
63. Lazcano-Ponce, E., Herrero, R., Munoz, N., Cruz, A., Shah, K. V., Alonso, P., Hernandez, P., Salmeron, J., and Hernandez, M. (2001). Epidemiology of HPV infection among Mexican women with normal cervical cytology. *Int J Cancer* **91**, 412–20.
64. Herrero, R., Hildesheim, A., Bratti, C., Sherman, M. E., Hutchinson, M., Morales, J., Balmaceda, I., Greenberg, M. D., Alfaro, M., Burk, R. D., Wacholder, S., Plummer, M., and Schiffman, M. (2000). Population-based study of human papillomavirus infection and cervical neoplasia in rural Costa Rica. *J Natl Cancer Inst* **92**, 464–74.
65. Ferreccio, C., Prado, R. B., Luzoro, A. V., Ampuero, S. L., Snijders, P. J., Meijer, C. J., Vaccarella, S. V., Jara, A. T., Puschel, K. I., Robles, S. C., Herrero, R., Franceschi, S. F., and Ojeda, J. M. (2004). Population-based prevalence and age distribution of human papillomavirus among women in Santiago, Chile. *Cancer Epidemiol Biomarkers Prev* **13,** 2271–6.
66. Molano, M., Posso, H., Weiderpass, E., van den Brule, A. J., Ronderos, M., Franceschi, S., Meijer, C. J., Arslan, A., and Munoz, N. (2002). Prevalence and determinants of HPV infection among Colombian women with normal cytology. *Br J Cancer* **87**, 324–33.
67. Matos, E., Loria, D., Amestoy, G. M., Herrera, L., Prince, M. A., Moreno, J., Krunfly, C., van den Brule, A. J., Meijer, C. J., Munoz, N., and Herrero, R. (2003). Prevalence of human papillomavirus infection among women in Concordia, Argentina: a population-based study. *Sex Transm Dis* **30**, 593–9.
68. Brestovac, B., Harnett, G. B., Smith, D. W., Shellam, G. R., and Frost, F. A. (2005). Human papillomavirus genotypes and their association with cervical neoplasia in a cohort of Western Australian women. *J Med Virol* **76**, 106–10.
69. Meekin, G. E., Sparrow, M. J., Fenwicke, R. J., and Tobias, M. (1992). Prevalence of genital human papillomavirus infection in Wellington women. *Genitourin Med* **68**, 228–32.
70. Thomas, J. O., Herrero, R., Omigbodun, A. A., Ojemakinde, K., Ajayi, I. O., Fawole, A., Oladepo, O., Smith, J. S., Arslan, A., Munoz, N., Snijders, P. J., Meijer, C. J., and Franceschi, S. (2004). Prevalence of papillomavirus infection in women in Ibadan, Nigeria: a population-based study. *Br J Cancer* **90**, 638–45.
71. Davidson, M., Schnitzer, P. G., Bulkow, L. R., Parkinson, A. J., Schloss, M. L., Fitzgerald, M. A., Knight, J. A., Murphy, C. M., Kiviat, N. B., Toomey, K. E., and et al. (1994). The prevalence of cervical infection with human papillomaviruses and cervical dysplasia in Alaska Native women. *J Infect Dis* **169**, 792–800.

Chapter 23

Epigenetic Targets in Cancer Epidemiology

Ramona G. Dumitrescu

Summary

Recently, it has been shown that epigenetic changes are involved in early stages of tumorigenesis, and they may trigger the genetic events leading to tumor development. In cancer epidemiology, there are several epigenetic alterations involved, such as DNA hypermethylation, DNA hypomethylation, and chromatin modifications with critical roles in the initiation and progression of human neoplasms. This chapter discusses the hypermethylation profiles of several tumor types, including bladder, brain, breast, colorectal, ovarian, prostate, and other cancers as well as DNA hypomethylation phenomena together with the chromatin modifications and their role in the complex mechanism of epigenetic silencing. Moreover, the involvement of environmental exposures in cancer susceptibility is addressed. In conclusion, these epigenetic changes are important characteristics of human neoplasia, and a better understanding of these modifications and the link between these changes for each tumor type will be important in early diagnosis of cancer and cancer prevention.

Key words: DNA hypermethylation, DNA hypomethylation, chromatin modifications, environment and cancer susceptibility.

1. Introduction

Traditionally, cancer is defined as a set of diseases, driven from genetic alterations such as mutations in oncogenes and tumor suppressor genes and chromosomal abnormalities. Lately, it has been revealed that the epigenetic changes are present during early stages of tumorigenesis *(1–3)*. It has been shown that cellular pathways such as DNA repair pathways are inactivated by the epigenetic silencing that leads to genomic instability and increased mutation rate *(2)*. Therefore, the early epigenetic changes of premalignant cells might determine the subsequent genetic alterations that lead

Mukesh Verma (ed.), *Methods in Molecular Biology, Cancer Epidemiology, vol. 471*
© 2009 Humana Press, a part of Springer Science + Business Media, Totowa, NJ
Book doi: 10.1007/978-1-59745-416-2

to cancer progression *(1)*. Epigenetic changes are alterations without the modification of the primary DNA sequence that are able to convey from one generation to another during cell division. It has been revealed that most of human cancers have epigenetic alterations that collaborate with genetic abnormalities during tumorigenesis. Epigenetic alterations in cancer cells involve simultaneously DNA hypermethylation of the gene promoter cytosine-phospho-guanosine (CpG) islands, global hypomethylation, and alterations of the chromatin structure. These changes may lead to abnormal silencing of critical genes to the initiation and progression of cancers, to activation of transcription of repeat elements, abnormal activation of genes, and deregulation of the chromosome replication process, which results in genomic instability. Moreover, DNA methylation and chromatin alterations contribute to the genes transcriptional activation or repression from early to late stages of tumor progression.

2. DNA Hypermethylation and Gene Silencing in Cancer

DNA methylation research in cancer has focused on CpG islands located in the gene's promoter regions, mostly because CpG islands' hypermethylation results in gene silencing both in normal key biological processes and in pathologic conditions, especially in cancer. In normal conditions, gene silencing is necessary in cellular differentiation, imprinting, and silencing of chromosome X in females. In human cancers, genes involved in cellular growth and differentiation, DNA repair, cell cycle regulation, hormone response, and cytokine signaling have been shown to present DNA hypermethylation at the CpG island promoter region, which has been linked to loss of transcriptional activity *(4)*. The first tumor suppressor gene shown to be silenced through DNA hypermethylation was *RB*, the gene associated with retinoblastoma *(5)*. Many other tumor suppressor genes, such as $p16^{INK4A}$, *VHL*, *MLH1*, *APC*, and *E-cadherin* are silenced in cancers through DNA methylation *(6)*. In population studies, DNA hypermethylation can be used as biomarker in cancer diagnosis either for early detection or disease classification *(7)*. Although promoter hypermethylation of certain key genes occurs in most cancer types, each tumor type presents a different pattern of genes with DNA hypermethylation, so different cancer types will be analyzed for different genes CpG island promoter hypermethylation markers.

2.1. Bladder Cancer

Recently, it has been shown that the methylation of $p16^{INK4A}$ and *p14(ARF)* is significantly higher in patients with invasive tumors than in superficial bladder cancer patients, suggesting that this methylation change can be a useful biomarker for the pathologic

stage, clinical outcome, and prognosis of patients with bladder cancer *(8)*. Moreover, six Wnt-antagonist genes' (*sFRP-1*, *sFRP-2*, *sFRP-4*, *sFRP-5*, *Wif-1*, and *Dkk-3*) hypermethylation have been shown to be involved in bladder cancer tumorigenesis, and the panel of genes can serve as an epigenetic biomarker for bladder cancers *(9)*. In noninvasive bladder cancers, it has been identified that DNA methylation of six genes (*SOCS-1*, *STAT-1*, *BCL-2*, *DAPK*, *TIMP-3*, and *E-cadherin*) was associated with bladder tumor recurrence *(10)*. Another study showed that hypermethylation at the promoter region of *RASSF1A*, *E-cadherin*, *TNFSR25*, *EDNRB*, and *APC* genes is associated with progression risk in bladder cancer *(11)*.

2.2. Brain Cancer

In astrocytoma tissues, *RASSF1A*, *p73*, *AR*, *MGMT*, *CDH1*, *OCT6*, *MT1A*, *WT1*, and *IRF7* genes were frequently hypermethylated, suggesting that these genes may be used for the development of prognosis or diagnosis markers for astrocytoma *(12)*. Another study showed that hypermethylation of *MGMT* and *p14(ARF)* is involved in the progression of astrocytoma *(13)*. Other two genes, *LATS1* and *LATS2*, were found to be hypermethylated and down-regulated in astrocytoma *(14)*. In medulloblastoma, several studies have identified tumor suppressor genes such as *RASSF1A*, *CASP8*, and *HIC1* that are inactivated through DNA hypermethylation, leading to gene silencing *(15)*. Furthermore, aberrant hypermethylation on *DAPK1* promoter has been detected, even in peripheral blood samples from patients with brain metastases *(16)*. The detection of this epigenetic alteration in blood could be a potential tumor prognosis marker. Moreover, methylation profiling has revealed a panel of genes that are target to DNA hypermethylation, leading to gene silencing in gliomas *(17)*. Recently, it has been shown that *hMLH1* promoter hypermethylation is associated with the development and progression of gliomas, and this is an early event in the tumorigenesis process *(18)*.

2.3. Breast Cancer

In breast carcinogenesis, DNA hypermethylation is involved in all six capabilities necessary for a cell to become malignant, according to Hanahan and Weinberg *(19)*. The six capabilities are limitless replicative potential, with the following genes found to be hypermethylated: $p16^{INK4A}$, *RIZ1*, *cyclin D2*, *RAR*β, *BRCA1*, *p53, 14-3-3*σ, and *GSTP1*; self-sufficiency in growth signals, when DNA methylation of the promoter region of genes such as *ER*α*, ER*β, *PR*, *SOCS1*, *RASSF1A*, and *SYK* plays a role in these genes' inactivation; insensitivity to growth-inhibitory signals, with genes such as *HIN-1*, *NES1* down-regulated through DNA hypermethylation; evasion of programmed cell death, with *HOXA5*, *DAPK*, *Twist*, *TMS1*, *GPC3*, *FHIT* genes' hypermethylation involved in this function; and sustained

angiogenesis, with hypermethylation of *maspin* and *THBS1* genes implicated in this process, and tissue invasion and metastasis resulting at least in part by *E-Cadherin*, *CDH13*, *APC, prostasin*, *TIMP-3*, and *BCSG1* genes' silencing through DNA hypermethylation *(20)*. In addition, DNA hypermethylation of several genes, including *RASSF1A*[21,22], *APC*[21,23], *BRCA1*, *p16*^{INK4A}, and *14-3-3σ (24,25)*, *p14ARF*, *cyclin D2*, and *Slit2*[25] have been detected even in serum of breast cancer patients. Moreover, evidence of promoter hypermethylation of *p16*^{INK4A} also has been observed in histologically normal breast tissues (26). In cultured human mammary epithelial cells, subpopulations of these cells have been shown to present *p16*^{INK4A} hypermethylation and accumulate chromosomal abnormalities, leading to a premalignant phenotype *(26)*.

2.4. Cervical Cancer

Human papillomavirus (HPV) is considered causative agent in the etiology of cervical cancer and among different types of this virus, HPV18 has been associated with more aggressive forms of cervical neoplasia. Recently, HPV18 L1 methylation has been observed mostly in carcinomas; therefore, L1 methylation may constitute a epigenetic marker for the neoplastic progression *(27)*. A study of several gynecologic cancers has found that there is a distinct panel of hypermethylated genes (*CDH1*, *DAPK*, *MGMT*, and *MINT2*) in cervical cancer and that the hypermethylation status of several of these genes may be useful biomarkers for tumor classification and disease stratification *(28)*. In cervical cancers, it has been shown that *MYOD1* gene promoter hypermethylation detected in the serum of cancer patients may be a potential prognostic marker for metastasis or recurrence *(29)*. Moreover, *cyclin A1* hypermethylation has been found frequently in cervical cancers, and it has been associated with the invasive phenotype, suggesting its potential as a tumor marker for early diagnosis of invasive cervical cancer *(30)*. Lately, it has been suggested that DNA hypermethylation of *COX-2* gene may be a possible prognostic marker in early stage cervical cancer *(31)*.

2.5. Colorectal Cancer

In colorectal cancers, methylation of *HLTF* and *HPP1/TPEF* has been associated with metastatic disease and tumor stage, and the detection of these changes in serum from colorectal cancer patients may be useful for the prediction of the disease outcome *(32)*. *MLH1* promoter hypermethylation is involved in silencing the *MLH1* gene expression and preventing normal activation of the DNA repair gene, leading to genomic instability and cell proliferation with role in colorectal cancer formation *(33)*. Moreover, DNA hypermethylation of *SFRP2* gene *(34)* has been detected in stools from patients with colorectal cancer and even in precancerous lesions. In addition, it has been suggested that *DAPK* promoter methylation is an early event in colorectal carcinogenesis *(35)*.

2.6. Lung Cancer

RASSF1A hypermethylation has been considered a biomarker for lung cancer, because it was found in bronchial aspirates collected from patients diagnosed with small and nonsmall cell lung carcinoma but not in patients diagnosed with nonneoplastic lung disease *(36)*. Hypermethylation of *ASC/TMS1* gene in sputum has been shown to indicate the recurrence of lung cancer in patients, with resection for early stage of the disease being a marker for late-stage lung cancer *(37)*. In non-small cell lung carcinoma, $p16^{INK4A}$ and *CDH13* hypermethylation also has been observed in patient serum, suggesting that this change can be a promising tool for non invasive early detection of lung cancer *(38)*. Moreover, promoter methylation of several genes, including $p16^{INK4A}$, *MGMT*, *DAPK*, *RASSF1A*, *PAX5*β, and *GATA5* in sputum from participants in a nested case-control study of incident lung cancer cases from a high-risk cohort has been shown to precede lung cancer incidence, therefore, promoter hypermethylation in sputum can be used as a molecular marker for identifying susceptible individuals to development of lung cancer *(39)*.

2.7. Prostate Cancer

Several frequently hypermethylated genes have been considered as potential prostate cancer-specific methylation markers. These genes are hormone response genes (*AR*, *ER*α, and *ER*β), cell cycle control genes (*CDKN2A*, with hypermethylation especially at the exon 2 of the gene), tumor invasion and tumor architecture genes (*E-cadherin*, *CD44*, *H-cadherin*, *APC*, *CAV1*, *LAMA3*, *LAMB3*, and *LAMC2*), DNA damage repair genes (*GSTP1* and *MGMT*), and genes such as *RASSF1 (40)*. It also has been shown that *PDLIM4* repression by hypermethylation may be a biomarker for prostate cancer *(41)*. *RAR*β*2* methylation levels can be associated with higher tumor stage *(42)*. Hypermethylation of *GSTP1* gene and a combined methylation profile of *APC* and *cyclin D2* in prostate cancer patients have been shown to be significant predictors for the time to relapse of prostate cancer *(43)*. *GSTP1* hypermethylation also can be detected in serum *(44–46)*, plasma *(44,47)*, urine and ejaculates *(44)* from prostate cancer patients so that this epigenetic change may be used in prostate cancer diagnosis and prognosis. It has been observed that *14-3-3*γ promoter methylation is frequently found in both benign and malignant prostate lesions *(48)*. Additionally, hypermethylation of tumor suppressor genes like *AR*, and *14-3-3*γ can be detected in serum of prostate cancer patients and even in serum of healthy controls *(46)*. Recently, it has been shown that *RASSF1A* methylation is present in both premalignant and benign epithelia, suggesting that this epigenetic change may be an effective marker of early prostate cancer *(49)*.

2.8. Ovarian Cancer

The incidence and mortality rate from ovarian cancer is almost the same, and due to less survival time, there is a need to identify markers for ovarian cancer. It has been suggested that promoter hypermethylation may be an alternative to mutation to inactivate tumor suppressor genes such as *MGMT*, *CDH1*, *RAR*β, and *SYK*, and this also can be an early event in the ovarian tumorigenesis *(50)*. *FHIT* hypermethylation also has been frequently observed in ovarian cancers, leading to gene silencing *(51)*. Furthermore, repetitive rDNA genes hypermethylation was found in ovarian tumors and rDNA methylation levels were associated with patients survival, suggesting that rDNA methylation may be used as a prognostic marker in ovarian cancer *(52)*. In addition, *IGFBP-3* promoter methylation has been associated with the disease progression and death in early-stage ovarian cancer *(53)*. Lately, it has been observed that ovarian cancers present a specific DNA hypermethylation profile of the genes *BRCA1*, *p14*, *p15*, *RIZ1*, and *TMS1 (54)*, which may be important for tumor detection and management.

2.9. Other Cancers

Hypermethylation of genes such as *TIMP3 (55)*, *ER*α *(56)*, and *RASSF1A*(57) has been shown to be involved in melanoma development and progression. Aberrant methylation of *RASSF1A*, $p16^{INK4}$, *O6-MGMT*, *RAR*β, and *hMLH1* is associated with tumor stage and histologic characteristics in pancreatic endocrine neoplasms *(58)*. In hepatocarcinoma, the aberrant methylation is an early and frequent event and the methylation of *E-cadherin* or *GSTP1* might serve as a potential prognostic marker for hepatocarcinoma patients *(59)*. In addition, *RASSF1A* methylation could be used as a marker for treatment outcome in hepatoblastoma *(60)*. Esophageal cancers exhibit CpG island hypermethylation of several genes, including $p14^{ARF}$, *p15*, $p16^{INK4}$, *FHIT*, $RAR\beta_2$, *APC*, *ER*, *MGMT*, *E-cadherin*, *TSLC1*, and *RASSF1A* genes, with a role in the early events and the progression of the esophageal carcinogenesis *(61)*. The profile of CpG island hypermethylation in hematologic malignancies is an epigenetic signature unique for each subtype of leukemia or lymphoma, with the following genes commonly found hypermethylated: *p15* and $p16^{INK4a}$ (specifically in acute myelogenous leukemia [AML] and acute lymphoblastic leukemia), and *MGMT*, *RARB*2, *CRBP1*, *SOCS-1*, *CDH1*, *DAPK*, and others *(62)*.

3. DNA Hypomethylation in Cancer

DNA hypomethylation can occur at the CpG dinucleotides at individual genes, at the global level, in the repeated DNA sequences, or in single copy genes *(63)*. Different cancer types show dif-

ferent levels of global hypomethylation (10–30% reduction in colorectal cancers and up to 50% reduction in breast cancers), and some tumor types do not exhibit significant reductions of the total level of 5-methyl cytosine (chronic myelogenous leukemia and AML) *(63)*. Moreover, the change in global hypomethylation levels seems to be an early event in tumorigenesis of breast, colon cancers and CLL. DNA hypomethylation in tumors also can be found in repeated DNA sequences such as simple repeats-satellite DNA and long and short interspersed repeats (LINE and SINE) *(63,64)*. Furthermore, in cancer, several growth-promoting genes are activated through hypomethylation, for example. *HRas*, *cyclin D2*, and *maspin* in gastric cancer, and carbonic anhydrase IX in renal-cell cancer. Moreover, human papilloma virus HPV16 is activated by hypomethylation in cervical cancers *(6)*. Therefore, DNA hypomethylation may serve as a useful diagnostic marker for human cancers.

4. Chromatin Changes and Their Role in Gene Silencing

DNA hypermethylation at the promoter region is not sufficient to induce the repression of the transcription and inactivation; it is necessary for it to be accompanied by a change of the chromatin structure in the nucleus. The histone modifications play a critical role in the maintenance of the chromatin structure. The most important changes of the histones are methylation and acetylation of the core histone proteins. Histone methylation is regulated by specific enzymes, namely, methyltransferases and demethylases, which play a role in maintaining the chromatin structure opened for the gene transcription (methylation of lysine 4 of histone 3) or condensed for gene repression (methylation of lysine 9 or lysine 27) *(65)*.

Histone acetyltransferases have an important role in maintaining the organization of the chromatin for gene transcription, whereas histone deacetylases are implicated in condensation of the chromatin and then repression of the transcription *(65)*.

Moreover, it has been shown that DNA methylation and the histone deacetylation, can be linked by the methyl-CpG-binding protein MeCP2, leading to gene transcription repression *(66)*. The connection between DNA methylation and histone modifications in several genes regulation has been observed in different tumor types, including prostate cancer, hepatocellular carcinomas, and other cancers *(40,67,68)*. Furthermore, it has been suggested that DNA methylation results from the interaction between methyl-CpG-binding domain-containing protein MBD2 and the nucleosome remodeling complex NuRD *(69)*.

Brahma (Brm), the catalytic component of the chromatin remodeling complex SWI/SNF has been associated with MeCP2 and with gene repression 70. Therefore, it has been recognized that DNA cytosine methylation, histone modifications and chromatid remodeling are interconnected, and any change in this cycle may lead to silencing of cancer genes *(3)*.

5. Epigenetics, Environmental Exposure, and Cancer Susceptibility

Lately, it has been indicated that environmental factors such as xenobiotic chemicals, nutritional supplements, behavior and reproductive factors, and low-doses radiation can lead to epigenetic alterations, and these changes may influence the risk of developing cancer *(71)*. Moreover, there is evidence that prenatal or early postnatal exposure to several environmental factors results in cancer susceptibility later in the adult life. It is believed that the mechanism by which the environmental exposures early in the development lead to disease phenotypes is through alterations of the epigenome. There are two sets of genes, the imprinted genes and genes with metastable epialleles that may link the early environmental exposure to later disease phenotypes. The imprinted genes are often clustered, and their expression is regulated by imprinted control regions so that if these regions are disturbed by genetic or epigenetic alterations, many genes are affected. There are several human genetic diseases associated with imprinting defects, including Beckwith-Wiedemann syndrome, Prader-Willi and Angelman syndromes, and human cancers *(72)*. Several human tumors such as Wilms' tumor, neuroblastoma, rhabdomyosarcoma, acute myeloblastic leukemia, and sporadic osteosarcoma are associated with the loss of a particular parental chromosome, indicating the involvement of imprinted genes *(72)*.

In pediatric Wilms tumors, loss of imprinting at the *IGF2* locus has been associated with aberrant methylation of the *H19* gene, which has a CpG island in the promoter that is normally methylated on the paternal allele and unmethylated on the maternal allele. Moreover, deregulation of *IGF2* imprinting is implicated in the development of several human cancers, including bladder, breast, cervical, choriocarcinoma, colorectal, glioma, leukemia-acute myeloid, lung, ovarian, prostate. and uterine *(72)*. Other imprinted genes, such as *ARHI* and *ZAC1*, have a role of tumor suppressor genes in breast and ovarian carcinoma *(73)*.

Metastable epialleles are represented by identical alleles that are expressed differently due to epigenetic changes during early embryogenesis *(74)*. There are several metastable epialleles (A^{vy}, $Axin^{Fu}$, and $Cabp^{IAP}$) well studied in mice, and the identification

and characterization of additional metastable epialleles in mice and in humans will help us to better understand how epigenetic changes contribute to disease susceptibility.

In conclusion, epigenetic changes are important characteristics of human neoplasia, and a better understanding of these modifications and the link between these changes for each tumor type will shed be valuable in early diagnosis of cancer and therefore of cancer prevention.

Acknowledgments

I thank Dr. Marian Catalin for reviewing this manuscript. This work was supported by the DOD grant Breast Center of Excellence-DAMD17-03-1-0446 (to Peter Shields).

References

1. Baylin, S.B. and Ohm, J.E. (2006) Epigenetic gene silencing in cancer–a mechanism for early oncogenic pathway addiction? *Nat. Rev. Cancer* **6**, 107– 116.
2. Jacinto, F.V. and Esteller, M. (2007) Mutator pathways unleashed by epigenetic silencing in human cancer. *Mutagenesis* **22**, 247–253
3. Jones, P.A. and Baylin, S.B. (2007) The epigenomics of cancer. *Cell* **128**, 683–692
4. Esteller, M. (2005) Aberrant DNA methylation as a cancer-inducing mechanism. *Annu. Rev. Pharmacol. Toxicol.* **45**, 629– 656.
5. Greger, V. et al. (1989) Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma. *Hum. Genet.* **83**,155– 158.
6. Feinberg, A.P. (2007) Phenotypic plasticity and the epigenetics of human disease. *Nature* **447,** 433– 440.
7. Verma, M. and Manne, U. (2006) Genetic and epigenetic biomarkers in cancer diagnosis and identifying high risk populations. *Crit Rev. Oncol. Hematol.* **60**, 9– 18.
8. Kawamoto, K. et al. (2006) p16INK4a and p14ARF methylation as a potential biomarker for human bladder cancer. *Biochem. Biophys. Res. Commun.* **339**, 790–796.
9. Urakami, S. et al. (2006) Wnt antagonist family genes as biomarkers for diagnosis, staging, and prognosis of renal cell carcinoma using tumor and serum DNA. *Clin. Cancer Res.* **12**, 6989–6997.
10. Friedrich, M.G. et al. (2007) DNA methylation on urinalysis and as a prognostic marker in urothelial cancer of the bladder. *Urologe A* **46,** 761–768.
11. Yates, D.R. et al. (2007) Promoter hypermethylation identifies progression risk in bladder cancer. *Clin. Cancer Res.* **13,** 2046–2053.
12. Yu, J. et al. (2004) Methylation profiles of thirty four promoter-CpG islands and concordant methylation behaviours of sixteen genes that may contribute to carcinogenesis of astrocytoma. *BMC Cancer* **4,** 65.
13. Watanabe, T. et al. (2007) Aberrant hypermethylation of p14ARF and O6-methylguanine-DNA methyltransferase genes in astrocytoma progression. *Brain Pathol.* **17,** 5–10.
14. Jiang, Z. et al. (2006) Promoter hypermethylation-mediated down-regulation of LATS1 and LATS2 in human astrocytoma. *Neurosci. Res.* **56,** 450–458.
15. Lindsey, J.C. et al. (2005) Epigenetic events in medulloblastoma development. *Neurosurg. Focus* **19,** E10.
16. Martinez-Glez, V. et al. (2007) DAPK1 promoter hypermethylaiton in brain metastases and peripheral blood. *Neoplasma* **54,** 123–126.
17. Kim, T.Y. et al. (2006) Epigenomic profiling reveals novel and frequent targets of aberrant DNA methylation-mediated silencing in malignant glioma. *Cancer Res.* **66,** 7490–7501.

18. Gomori, E. et al. (2007) Epigenetic inactivation of the hMLH1 gene in progression of gliomas. *Diagn. Mol. Pathol.* **16,** 104–107.
19. Hanahan, D. and Weinberg, R.A. (2000) The hallmarks of cancer. *Cell* 100, 57–70.
20. Widschwendter, M. and Jones, P.A. (2002) DNA methylation and breast carcinogenesis. *Oncogene* **21,** 5462–5482.
21. Muller, H.M. et al. (2004) Prognostic DNA methylation marker in serum of cancer patients. *Ann. N Y Acad. Sci.* **1022,** 44–49.
22. Taback, B. et al. (2006) Epigenetic analysis of body fluids and tumor tissues: application of a comprehensive molecular assessment for early-stage breast cancer patients. *Ann. N Y Acad. Sci.* **1075,** 211–221.
23. Zhang, J.J. et al. (2007) [Detection and significance of APC gene promoter hypermethylation in serum of breast cancer patients]. *Ai. Zheng* **26,** 44–47.
24. Jing, F. et al. (2007) Hypermethylation of tumor suppressor genes BRCA1, p16 and 14-3-3sigma in serum of sporadic breast cancer patients. *Onkologie* 30, 14–19.
25. Sharma, G. et al. (2007) Promoter hypermethylation of p16INK4A, p14ARF, CyclinD2 and Slit2 in serum and tumor DNA from breast cancer patients. *Life Sci.* **80,** 1873–1881.
26. Holst, C.R. et al. (2003) Methylation of p16(INK4a) promoters occurs in vivo in histologically normal human mammary epithelia. *Cancer Res.* **63,** 1596–1601.
27. Turan, T. et al. (2006) Methylation of the human papillomavirus-18 L1 gene: a biomarker of neoplastic progression? *Virology* **349,** 175–183.
28. Yang, H.J. et al. (2006) Differential DNA methylation profiles in gynecological cancers and correlation with clinico-pathological data. *BMC Cancer* **6,** 212.
29. Widschwendter, A. et al. (2004) DNA methylation in serum and tumors of cervical cancer patients. *Clin. Cancer Res.* **10,** 565–571.
30. Kitkumthorn, N. et al. (2006) Cyclin A1 promoter hypermethylation in human papillomavirus-associated cervical cancer. *BMC Cancer* 6, 55.
31. Jo, H. et al. (2007) Hypermethylation of the COX-2 gene is a potential prognostic marker for cervical cancer. *J. Obstet. Gynaecol. Res.* **33,** 236–241.
32. Wallner, M. et al. (2006) Methylation of serum DNA is an independent prognostic marker in colorectal cancer. *Clin. Cancer Res.* **12,** 7347–7352.
33. Niv, Y. (2007) Microsatellite instability and MLH1 promoter hypermethylation in colorectal cancer. *World J. Gastroenterol.* **13,** 1767–1769.
34. Huang, Z. et al. (2007) Hypermethylation of SFRP2 as a potential marker for stool-based detection of colorectal cancer and precancerous lesions. *Dig. Dis. Sci.* **52,** 2287–2291.
35. Mittag, F. et al. (2006) DAPK promotor methylation is an early event in colorectal carcinogenesis. *Cancer Lett.* **240,** 69–75.
36. Grote, H.J. et al. (2006) Methylation of RAS association domain family protein 1A as a biomarker of lung cancer. *Cancer* **108,** 129–134.
37. Machida, E.O. et al. (2006) Hypermethylation of ASC/TMS1 is a sputum marker for late-stage lung cancer. *Cancer Res.* **66,** 6210–6218.
38. Ulivi, P. et al. (2006) p16INK4A and CDH13 hypermethylation in tumor and serum of non-small cell lung cancer patients. *J. Cell Physiol* **206,** 611–615.
39. Belinsky, S.A. et al. (2006) Promoter hypermethylation of multiple genes in sputum precedes lung cancer incidence in a high-risk cohort. *Cancer Res.* **66,** 3338–3344.
40. Li, L.C. (2007) Epigenetics of prostate cancer. *Front Biosci.* **12,** 3377–3397.
41. Vanaja, D.K. et al. (2006) PDLIM4 repression by hypermethylation as a potential biomarker for prostate cancer. *Clin. Cancer Res.* **12,** 1128–1136.
42. Jeronimo, C. et al. (2004) Quantitative RARbeta2 hypermethylation: a promising prostate cancer marker. *Clin. Cancer Res.* **10,** 4010–4014.
43. Rosenbaum, E. et al. (2005) Promoter hypermethylation as an independent prognostic factor for relapse in patients with prostate cancer following radical prostatectomy. *Clin. Cancer Res.* **11,** 8321–8325.
44. Henrique, R. and Jeronimo, C. (2004) Molecular detection of prostate cancer: a role for GSTP1 hypermethylation. *Eur. Urol.* **46,** 660–669.
45. Bastian, P.J. et al. (2005) Preoperative serum DNA GSTP1 CpG island hypermethylation and the risk of early prostate-specific antigen recurrence following radical prostatectomy. *Clin. Cancer Res.* **11,** 4037–4043.
46. Reibenwein, J. et al. (2007) Promoter hypermethylation of GSTP1, AR, and 14-3-3sigma in serum of prostate cancer patients and its clinical relevance. *Prostate* **67,** 427–432.

47. Papadopoulou, E. et al. (2006) Cell-free DNA and RNA in plasma as a new molecular marker for prostate and breast cancer. *Ann. N. Y. Acad. Sci.* **1075,** 235–243.
48. Henrique, R. et al. (2005) Frequent 14-3-3 sigma promoter methylation in benign and malignant prostate lesions. *DNA Cell Biol.* **24,** 264–269.
49. Aitchison, A. et al. (2007) RASSF1A promoter methylation is frequently detected in both pre-malignant and non-malignant microdissected prostatic epithelial tissues. *Prostate* **67,** 638–644.
50. Dhillon, V.S. et al.(2004) Promoter hypermethylation of MGMT, CDH1, RAR-beta and SYK tumour suppressor genes in granulosa cell tumours (GCTs) of ovarian origin. *Br. J. Cancer* **90,** 874–881.
51. Hong, F.Z. et al. (2005) Hypermethylation of fragile histidine triad gene and 3p14 allelic deletion in ovarian carcinomas. *Zhonghua Bing. Li Xue. Za Zhi.* **34**, 257–261.
52. Chan, M.W. et al. (2005) Hypermethylation of 18S and 28S ribosomal DNAs predicts progression-free survival in patients with ovarian cancer. *Clin. Cancer Res.* **11,** 7376–7383.
53. Wiley, A. et al. (2006) Methylation of the insulin-like growth factor binding protein-3 gene and prognosis of epithelial ovarian cancer. *Int. J. Gynecol. Cancer* **16,** 210–218.
54. Yang, H.J. et al. (2006) Differential DNA methylation profiles in gynecological cancers and correlation with clinico-pathological data. *BMC Cancer* **6,** 212.
55. van'd., V. et al. (2003) Expression profiling reveals that methylation of TIMP3 is involved in uveal melanoma development. *Int. J. Cancer* **106,** 472–479.
56. Mori, T. et al. (2006) Estrogen receptor-alpha methylation predicts melanoma progression. *Cancer Res.* **66,** 6692–6698.
57. Maat, W. et al. (2007) Epigenetic inactivation of RASSF1a in uveal melanoma. *Invest. Ophthalmol. Vis. Sci.* **48,** 486–490.
58. House, M. G. et al. (2003) Aberrant hypermethylation of tumor suppressor genes in pancreatic endocrine neoplasms. *Ann. Surg.* **238,** 423–431.
59. Lee, S. et al. (2003) Aberrant CpG island hypermethylation along multistep hepatocarcinogenesis. *Am. J. Pathol.* **163,** 1371–1378.
60. Sugawara, W. et al. (2007) Promoter hypermethylation of the RASSF1A gene predicts the poor outcome of patients with hepatoblastoma. *Pediatr. Blood Cancer* **49,** 240–249.
61. Wu, D.L. et al. (2006) Methylation in esophageal carcinogenesis. *World J. Gastroenterol.* **12,** 6933–6940.
62. Esteller, M. (2003) Profiling aberrant DNA methylation in hematologic neoplasms: a view from the tip of the iceberg. *Clin. Immunol.* **109,** 80–88.
63. Wilson, A.S. et al. (2007) DNA hypomethylation and human diseases. *Biochim. Biophys. Acta* **1775,** 138–162.
64. Ehrlich, M. (2002) DNA methylation in cancer: too much, but also too little. *Oncogene* **21,** 5400–5413.
65. Brock, M.V. et al. (2007) Cancer as a manifestation of aberrant chromatin structure. *Cancer J.* **13**, 3–8.
66. Nan, X. et al. (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393,** 386–389.
67. Kondo, Y. et al. (2007) Alterations of DNA methylation and histone modifications contribute to gene silencing in hepatocellular carcinomas. *Hepatol. Res.*.
68. Vincent, A. et al. (2007) Epigenetic regulation (DNA methylation, histone modifications) of the 11p15 mucin genes (MUC2, MUC5AC, MUC5B, MUC6) in epithelial cancer cells. *Oncogene*.
69. Zhang, Y. et al. (1999) Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev.* **13,** 1924–1935.
70. Harikrishnan, K.N. et al. (2005) Brahma links the SWI/SNF chromatin-remodeling complex with MeCP2-dependent transcriptional silencing. *Nat. Genet.* **37,** 254–264.
71. Jirtle, R.L. and Skinner, M.K. (2007) Environmental epigenomics and disease susceptibility. *Nat. Rev. Genet.* **8,** 253–262.
72. Falls, J.G. et al. (1999) Genomic imprinting: implications for human disease. *Am. J. Pathol.* **154,** 635-647.
73. Weidman, J.R. et al. (2007) Cancer susceptibility: epigenetic manifestation of environmental exposures. *Cancer J.* **13,** 9–16.
74. Rakyan, V.K. et al. (2002) Metastable epialleles in mammals. *Trends Genet.* **18,** 348–351.

Chapter 24

Epidemiology of Lung Cancer Prognosis: Quantity and Quality of Life

Ping Yang

Summary

Primary lung cancer is very heterogeneous in its clinical presentation, histopathology, and treatment response; and like other diseases, the prognosis consists of two essential facets: survival and quality of life (QOL). Lung cancer survival is mostly determined by disease stage and treatment modality, and the 5-year survival rate has been in a plateau of 15% for three decades. QOL is focused on life aspects that are affected by health conditions and medical interventions; the balance of physical functioning and suffering from treatment side effects has long been a concern of care providers as well as patients. Obviously needed are easily measurable biologic markers to stratify patients before treatment for optimal results in survival and QOL and to monitor treatment responses and toxicities. Targeted therapies toward the mechanisms of tumor development, growth, and metastasis are promising and actively translated into clinical practice. Long-term lung cancer (LTLC) survivors are people who are alive 5 years after the diagnosis. Knowledge about the health and QOL in LTLC survivors is limited because outcome research in lung cancer has been focused mainly on short-term survival. The independent or combined effects of lung cancer treatment, aging, smoking and drinking, comorbid conditions, and psychosocial factors likely cause late effects, including organ malfunction, chronic fatigue, pain, or premature death among lung cancer survivors. New knowledge to be gained should help lung cancer survivors, their healthcare providers, and their caregivers by providing evidence for establishing clinical recommendations to enhance their long-term survival and health-related QOL.

Key words: Non-small cell lung cancer, small cell lung cancer, survival, quality of life, pharmacogenomics, gene expression profiling, chemotherapy, radiation therapy, toxicity.

1. Introduction

Lung cancer survivorship has two critical attributes: survival time or quantity and quality of life (QOL). The unparalleled perception regarding the importance of QOL versus survival after a lung cancer diagnosis was not appreciated by the medical

Mukesh Verma (ed.), *Methods in Molecular Biology, Cancer Epidemiology, vol. 471*
© 2009 Humana Press, a part of Springer Science + Business Media, Totowa, NJ
Book doi: 10.1007/978-1-59745-416-2

research community until recently, mainly due to the leading battle to prolong patients' lives and the devastating demand to find a cure. After decades of efforts focusing on reducing lung cancer incidence and mortality, we are now challenged by the lack of understanding of the health conditions and QOL among people who survived lung cancer. In this chapter, current yet limited knowledge is summarized in three major areas with varying levels of details that are available in the published literature: clinical epidemiology, genomic application toward patient-specific care, and psychosocial-behavioral characteristics of lung cancer survivors.

2. Clinical Epidemiology of Lung Cancer Survival

2.1. Overview

Primary lung cancer is very heterogeneous in its clinical presentation, histopathology, and treatment response. Conventionally, lung cancer has been divided into non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). For both groups, cancer stage is the most significant predictor of survival *(1,2)* as illustrated in **Fig. 24.1**. Five-year survival rates were comparable between two large clinical patient data resources that were from different institutions at different time frames: one from MD Anderson Cancer Center (diagnosed 1975–1988) and one from Mayo Clinic (diagnosed 1997–2002), indicating little improvement in NSCLC survival in the past 30 years. In SCLC, the 5-year

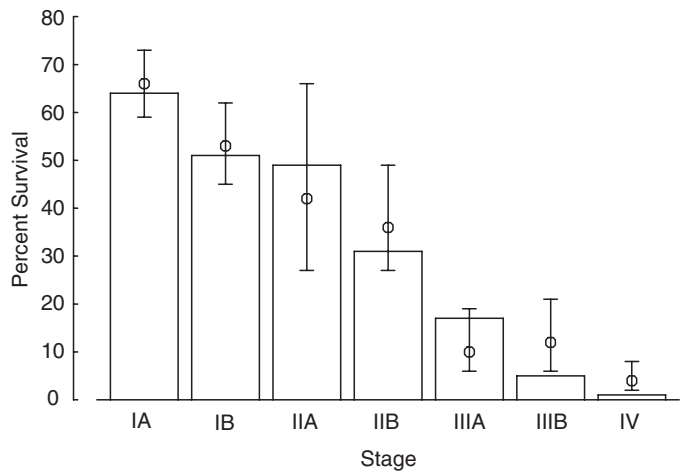


Fig. 24.1. Five-year survival rates and 99% confidence intervals of Mayo Clinic NSCLC patients compared with MD Anderson Cancer Center's Patients. Open bars are average survival rates for each disease stage based on Mountain's series of MD Anderson Cancer Center; circles with lines are estimated survival rates and 99% confidence intervals of Mayo Clinic's series. (Adapted from Yang et al. (2005), reference 2)

survival rate ranges from approximately 25% for limited disease to 1–5% for extensive disease *(2–4)*.

Treatment modality, mostly dictated by lung cancer stage and patient's performance status, directly determines disease survival. For NSCLC, surgical resection is the standard treatment for stage I–II disease; patients with IB or II disease are now being offered adjuvant chemotherapy *(5)*. Some patients with a stage IIIA tumor are operable but often receive pre- or postoperative radiation, chemotherapy, or both. For locally advanced, inoperable stage IIIA tumors, radiation with chemotherapy remains the standard of care; for selected patients with IIIB (with pleural effusion) or IV disease, chemotherapy remains the standard treatment in conjunction with supportive care. In SCLC, the majority of the patients with limited disease are treated with single-agent or combination chemotherapy with radiation therapy, and prophylactic cranial irradiation (for complete responders) is commonly used. In both NSCLC and SCLC, systemic dissemination can occur as solitary metastasis, oligometastasis, and multiple metastases. Solitary metastasis constitutes the majority of long-term survivors in advanced stage lung cancer *(6)*.

At present, for any given chemo- or radiotherapy regimen, which is prescribed by the standard protocol with a fixed dosage, the response rate and toxicity range vary widely among patients. The disease-free survival time, even after adjusting for the well-established predictors, that is, disease stage, histology, performance status, and treatment, varies significantly. Obviously needed are easily measurable biologic markers to stratify patients before treatment for optimal results in survival and QOL and to monitor treatment responses and toxicities.

2.2. Lung Cancer Progression and Subsequent Cancers

Despite the generally accepted diagnostic criteria, differentiation between lung cancer recurrence in the lung and occurrence of a new primary lung cancer is often difficult, especially when the subsequent tumor arises after >2 years from the initial diagnosis with the same tumor histology *(7,8)*. Molecular biologic methods have not yet been proven to be a reliable tool to meet this challenge. Recognizing this uncertainty in differential diagnosis, the published literature suggests that a majority of NSCLC recurrences occur within 2 years after surgery and late recurrence may occur up to 10 years *(9)*. Recurrent disease has been reported in 4–5% of 5-year NSCLC survivors *(9–12)*; up to 10% of recurrences may be discovered beyond 5 years after initial curative therapy *(13)*. Two cohort studies of disease-free 5-year survivors of NSCLC have reported a subsequent recurrence rate of 2–3% per patient-year *(10,14)*. Clinical management and outcome of recurrent lung cancer have been reported, demonstrating benefit of various treatments over no treatment *(15,16)*. Molecular characteristics or risk behaviors associated with late recurrent tumors have not been well described.

2.3. Additional Lung Cancer Prognostic Predictors

More and more factors emerged as independent lung cancer survival predictors in the recent decade. The importance of pathologic markers such as tumor size, cell type, lymphatic and blood vessel invasion, rate of proliferation and ploidy, and extent of tumor necrosis is apparent but inconsistent. The rest of this chapter highlights current knowledge of selected survival predictors beyond disease stage, treatment, and patient gender *(17)*.

2.3.1. Tumor Cell Differentiation

In a study of 5,018 hospital-based patients and 712 population-based patients, tumor grade was found to be significantly associated with survival after adjusting for the effects of age, gender, smoking history, tumor stage, histologic cell type, and treatment modality. Patients with poorly/undifferentiated carcinoma had a 70% elevated risk of death compared with those with well-differentiated carcinoma. A 40% elevated risk was observed for patients with moderately differentiated carcinoma *(18)*.

2.3.2. Smoking Cessation

In a study of 5,229 patients with NSCLC and SCLC, the median survival time among never, former, and current smokers with NSCLC was 1.4, 1.3, and 1.1 years, respectively ($P < 0.01$). Female NSCLC patients had a significantly lower risk of mortality with a longer duration of smoking abstinence. Specifically, the relative risk per 10 years of smoking abstinence is 0.85, supporting a direct biologic effect of smoking on survival *(19)*.

2.3.3. Dietary Supplements

In the general population, approximately 40% of people take vitamin/mineral supplements regularly; whereas nearly 80% of cancer patients do so. Both clinical and laboratory data have shown that certain micronutrients effect the growth of malignant cells, i.e., vitamins and minerals may be modulators of tumor growth. In a study of more than 1,300 lung cancer patients, the use of vitamin/mineral supplements was found to be associated with improved survival among both NSCLC and SCLC patients *(20,21)*. Reduction in the death rate was 26% for NSCLC patients and 37% for SCLC patients, respectively.

3. Genomic and Biologic Markers: Advances and Applications in Lung Cancer Prognosis

Since the completion of the sequence of the whole human genome, the genomic approach in cancer research is playing a significant role in identifying diagnostic, prognostic, and therapeutic molecular markers. During the past two decades, there have been >1,000 studies on biologic markers, including biochemical, histologic, and molecular, in the hopes of identifying clinically useful prognostic markers and to assist in patient management *(22)*. However,

incomplete scientific conclusions can be drawn because most of these studies are limited to small numbers or selected subgroups of patients, diverse methodology and type of specimens used, variable histogolic subtype, short and/or incomplete follow-up, univariate analysis, and various endpoints. The majority of the investigators have only looked at overall survival, mostly up to 5 years, with rare emphasis on disease-free survival, relapse rate, relationship to treatment responses, or long-term outcomes. In addition, there are only a few tumor target markers and virtually no host-specific predictive measures to base an optimal choice of the type, dosage, intensity, and combinations of the available drug(s).

3.1. Tumor Molecular Markers

3.1.1. Gene Expression Profiling: DNA Microarray Technology

DNA microarray performs simultaneous interrogation of thousands of genes, which offers a unique opportunity to measure a tumor from multiple angles, which generally provides a more accurate measurement about biologic behaviors than any single cellular or molecular parameter. As a high-throughput tool at the molecular level, DNA microarray has clear advantages over traditional histologic examinations, and it has been widely used in cancer research to better predict clinical outcomes and potentially improve patient management. The molecular measurement is more objective and often detects the difference that routine pathology fails to detect. More importantly, the DNA microarray provides a closer look at gene activities in tumors and creates an opportunity to find therapeutic targets. Studies show that this new approach provides accurate tumor subclassification and outcome predictions such as tumor stage, metastatic status, and patient survival, and it offers some hope for individualized medicine *(23)*. However, growing evidence suggests that gene-based prediction is not stable and little is known about the prediction power of the gene expression profile compared with well-known clinical and pathologic predictors. Overall, most studies lack an independent validation. When conventional predictors of age, gender, stage, cell type, and histologic grade are considered collectively, the predictive advantage of the gene expression profile diminishes. As shown in **Fig. 24.2**, the gene panel derived from the DNA microarray achieved very good prediction for patient survival, but it did not outperform the prediction by the combination of five conventional variables (age, gender, stage, cell type, and tumor grade). This suggests that most prediction from the gene panel is reflected by tumor pathology information; thus, the enhanced predictive accuracy from the gene panel is limited at this point.

3.1.2. Protein Expression Profiling: Mass Spectrometry

Only in the recent 5 years, technology in matrix-assisted laser desorption-ionization time-of-flight mass spectrometry (MALDI-TOF MS) made it possible to profile proteins in tissues. An obvious advantage over RNA expression profiling is that deciphering protein expression patterns would be closer to

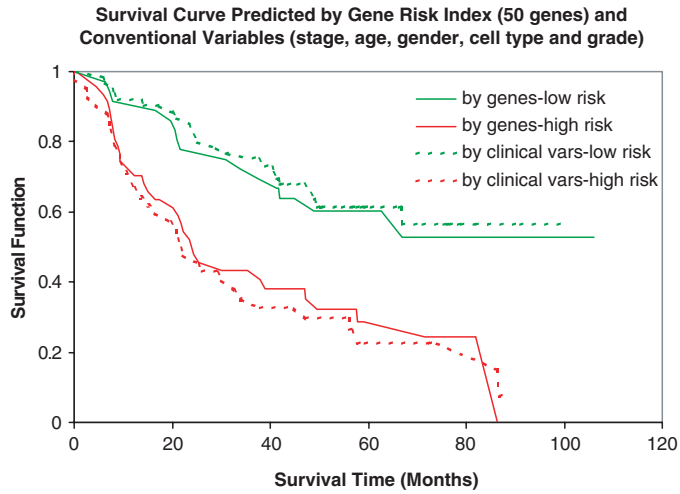


Fig. 24.2. Survivial curve predicted by Gene Risk Index (50 genes) and conventional variables (stage, age, gender, cell type and stage) *(23)*. Adapted from Sun and Yang (2006) CEPB.

revealing the biological function of cells, and then to determining the pathologic mechanisms of disease development and progression. Meanwhile, greater challenges have been encountered in analyzing and interpreting protein profiling due to the immense throughput of data and ultrasensitivity of signal detection. Pioneer researchers in applying protein profiling to predict lung cancer prognosis have reported a 15-protein peak-pattern *(24)* and a 25-signal proteomic signature *(25)* that can distinguish lung cancer patients with poor versus good outcome.

Genomic or proteomic signatures selected for predicting lung cancer recurrence and survival hold high hope to be used in clinical practice, particularly for those that have been vigorously validated in independent data sets. Like a clinical trial for a novel agent or a new treatment regimen before moving into the clinic, these markers need to be evaluated for their efficacy, demonstrating better than currently available predictors and ease of use. Clearly, a gene signature or panel that does not provide improved prediction independently from well-known predictors will less likely be adopted clinically.

3.2. Development of Targeted Therapy

3.2.1. Tumor Marker Targeted Therapy

An ultimate goal of searching for DNA marker panels, molecular pathway signatures, or genomic patterns is to discover biological targets for therapeutics. Many drugs that target molecular markers have been evaluated in randomized controlled trials, but most of them failed to produce positive results *(26)*. One of the important findings in cancer chemotherapy is the identification of somatic mutations in the tyrosine kinase domain of the epidermal growth factor receptor (EGFR) in NSCLC and a correlation with response to EGFR inhibitors *(27–29)*. EGFR gene

amplification is more prevalent in Western populations, whereas the amplification of the closely related HER2 gene, which also could have implications for the treatment of NSCLC, is more frequent in East Asian patients. Ethnicity may indicate different genetic backgrounds in common tumors that may influence clinical outcome and response to therapy *(30)*. Another important molecular marker for targeted therapy is the vascular endothelial growth factor (VEGF), which participates in the regulation of new vessel growth and promotion of immature vasculature survival *(31)*. Drugs that inhibit the VEGF have been developed and tested successfully in randomized controlled trials, with the caveat of severe side effects, including thrombocytopenia, hypertension, and neutropenia *(32)*. The combined use of targeted therapy and chemotherapy is very promising in both improved treatment response and reduced toxicity *(33)*.

3.2.2. Immunotherapy

The fundamental basis of immunotherapy is that the immune system can distinguish cancerous cells from normal cells, and the immune machinery can be used to destroy cancer cells. Three strategies have been taken *(34)*. The first strategy is nonspecific cellular immunotherapy, such as lymphokine-activated killer cells and interlukin-2 in treating renal cell carcinoma. The second is specific cellular immunotherapy that relies on but is not limited to tumor antigen-specific cytolytic lymphocytes and tumor infiltrating lymphocytes. The third strategy is therapeutic vaccination that incorporates tumor antigens with an adjuvant therapy for the recognition by the immune system. NSCLC has not been considered as an immunogenic cancer and presently there is no standard immunotherapy in clinical practice, although optimism prevails on the progress of evaluating novel immunotherapeutics *(35)*.

3.2.3. Gene Therapy

Improved molecular technology has made gene therapy an emerging and promising strategy for cancer treatment. Three experimental approaches have been used: 1) immunotherapy that applies genetically modified cells and viral particles to stimulate the immune system to kill cancer cells, 2) oncolytic virotherapy that causes cell death by replicating viral particles within cancer cells, 3) and gene transfer that causes cell death or low growth by introducing new gene(s) to cancer cells or surrounding tissue *(36)*. The promise of gene therapy is the identification of critical regulatory sequences that can selectively activate or inactivate the therapeutic genes in cancer cells; however, a major challenge is the development of the delivery of vector constructs that can reach the target effectively.

3.3. Host Genetic Predisposition in Treatment Response and Toxicity

Inherent and acquired drug resistance is a cause of chemotherapy failure, and pharmacogenomic studies have begun defining multiple gene variations responsible for varied drug metabolisms. Platinum-based drugs are the most commonly used in lung cancer

treatment, and a major obstacle for the clinical use of platinum drugs is the development of tumor resistance. Maintaining an effective antineoplastic level of the platinum drug in the sera and tumor for a prolonged period could potentially eliminate acquired resistance; however, such an approach increases the risk for toxicity and drug side effects. Cisplatin-induced ototoxicity, for example, is a result of drug-induced neuron and hair cell destruction and a major dose-limiting adverse drug response (ADR) *(37)*. Other platinum drug-related ADRs range from alopecia, neurotoxicity, myelotoxicity, to nephrotoxicity *(38,39)*. In contrast, emerging evidence suggests that mild-to-moderate levels of specific ADRs may be a direct gauge for accurate dosing *(40)*. An analysis of three randomized clinical trials *(39)* showed a significant survival benefit for NSCLC patients who had mild or severe chemotherapy-induced neutropenia compared with those who did not. Other studies have shown a similar relation between ADRs and treatment outcomes for various cancers, such as colorectal, breast, testicular, and ovarian *(41–44)*. The identification of patients most likely to benefit from chemotherapy through the incorporation of genomic information into treatment decisions may be a critical step in overcoming the barriers caused by ADRs *(45,46)*.

The glutathione metabolic pathway is directly involved in the detoxification or inactivation of platinum compounds; evidence supports the role of the glutathione pathway in acquired and inherited drug resistance through rapid drug detoxification or through drug activation bypassing, which adversely impacts the treatment outcome of lung cancer *(47)*. DNA repair is another critical mechanism of resistance to platinum-based chemotherapy. It is hypothesized that reduced DNA repair in tumor cells has a higher sensitivity to treatment and therefore has a better response and outcome after radio- and/or chemotherapy, whereas increased repair capability causes tumor resistance and worse response *(48)*. Clinical studies show that overexpression of excision repair cross-complementation group 1 (ERCC1) correlates with poor survival for gemcitabine/cisplatin-treated NSCLC patients, and the polymorphisms of ERCC1 or excision repair cross-complementation group 2 (ERCC2) are significantly associated with survival times for patients treated with platinum-based chemotherapy *(49–52)*.

Ionizing radiation is also a commonly used treatment modality for late stage lung cancer and acts on DNA, causing double-strand and single-strand breaks and base lesions, particularly double-strand DNA breaks *(53)*. The damage is repaired by at least two distinct pathways: Homologous recombination (HR) and nonhomologous end-joining (NHEJ). HR requires an undamaged template molecule that contains a homologous DNA sequence, generally from its sister chromatid; RAD51 and RAD52 proteins are involved in this pathway. For NHEJ, no

undamaged partner DNA homologs are needed for rejoining of DNA breaks; RAD50 and DNA-dependent protein kinase may participate in the NHEJ repair process *(54,55)*. Genetic defects in HR or NHEJ can cause impaired DNA replication and enhanced radiation sensitivity *(54)*.

A key question is whether biologic functions and mechanisms can be a clinically measurable determinant of a cancer patient's response to chemo/radiation therapy and survival. Specifically, although evidence from laboratory studies supports a significant role for the glutathione system and DNA repair pathway in antitumor drug metabolism and resistance, the exploration of their clinical relevance and implications in cancer patients continues to evolve. Regulation of the glutathione metabolic and transport systems has been one of the targets for optimizing efficacy as well as minimizing toxicities of many chemotherapeutic agents *(56–60)*. The premise for the clinical application of genomic polymorphic markers is most profound in identifying likely responsive patients and differentiating patients, for example, who could be sensitized with a glutathione and/or DNA repair modulator from those who should not be given the drug due to the predicted occurrence of severe side effects. Successful translation and application of these genetic polymorphic markers to treatment modulators or response predictors is an essential step toward individualized drug therapy.

4. Quality of Life: Psychosocial-Behavioral Characteristics among Lung Cancer Survivors

QOL is a subjective, dynamic, multidimensional measure encompassing all aspects that impact an individual's life *(61–63)*, and it is often exchangeable with the term health-related QOL (HRQOL). Although there is no consensus on the definition, QOL or HRQOL is more focused on life aspects that are affected by health conditions and medical interventions. Commonly used QOL instruments are well organized and summarized by Li et al. *(63)*: from general tools such as the World Health Organization Quality of Life Assessment *(62)* and Medical Outcome Study 36-Item Short-Form Health Survey (SF 36) *(64,65)* to cancer-specific measures such as the European Organization of Research and Treatment of Cancer (EORTC) Core Cancer questionnaire (QLQ C-30) *(66)*, and to lung cancer-specific scales such as the Functional Assessment of Cancer Therapy-Lung (FACT-L) version 3 *(67)* and Lung Cancer Symptom Scale (LCSS) *(68)*. A comprehensive review of QOL in lung cancer patients evaluated 50 instruments and identified the best tool to be the EORTC Quality

of Life Lung Cancer Questionnaire (EORTC-LC13) in conjunction with the QLQ-C30 *(69)*. LCSS and FACT-L are two additional instruments with good reliability and validity.

4.1. QOL Surrounding Lung Cancer Treatment

The balance of physical functioning and suffering from treatment side effects has long been a concern of patient care providers. Until two decades ago in the mid-1980s, supportive care plus combination chemotherapy (cisplatin and vinblastine) was not considered superior over supportive care only (palliative radiation, psychosocial support, analgesics, and nutritional support) for metastatic NSCLC because of the nonsignificant survival benefit with serious toxicity *(70)*. Ten years later, in a multicenter randomized phase III trial, for both QOL and survival, supportive care plus a different combination chemotherapy (carboplatin and etoposide) were shown superior to supportive care only *(71)*. In contrast, patients also have clear preference in value potential survival benefit and chemotherapy-induced toxicity *(72)*. Among 81 patients previously treated with platinum drugs for stage III/IV NSCLC, >50% of them would not choose chemotherapy for an estimated additional 3 months of survival. Some patients would not choose chemotherapy to avoid any interference with their QOL even for an estimated 24-month additional survival period, whereas others would take chemotherapy regardless of how long they would have to live as not to miss any opportunity to be cured. Lung cancer patients care about more than just survival and the majority of them would choose treatment that improves cancer symptoms.

Because surgical resection remains the treatment of choice for early stage NSCLC, prospective evaluation and preservation of long-term QOL after the surgery is imperative. Before the resection, patients already have lower QOL, mainly in physical and emotional functioning; their QOL was further impaired after the resection, particularly 3 to 6 months after operation *(63)*. More specifically, pulmonary resection is known to cause post-thoracotomy pain syndrome, commonly seen in approximately 50% of patients after thoracotomy. This chronic condition has been reported to last for 4 to 5 years in approximately 30% of patients *(73)*, with no data beyond 5 years after surgery. Another study of 224 patients pathologically diagnosed with NSCLC reports a strong effect of psychosocial factors on mortality one year after diagnosis. These factors include high need for sympathy and devotion, reserved personality, and either very low or very high ego strength *(74)*. Other known predictors considered in the study were disease stage, tumor cell type, and co-morbid conditions, disease stage was the only significant predictor.

Compared with NSCLC, SCLC is a fast growing tumor, and in general it is considered as a systemic disease. For numerous clinical trials of chemoradiation therapy combinations, QOL has

been measured, at the best, as secondary endpoints. In one of such phase III randomized trial of >500 patients, the trial arm (TEC) and the control arm (VEC) showed comparable overall QOL, function, and symptom scores *(75)*.

A key domain of QOL is symptoms. Lung cancer patients experience high symptom burden and distress, even among high-functioning patients *(76)*. Fatigue, pain, dysnea, anorexia, and cachexia are the most common symptoms, and they can be caused by lung cancer or side effects of treatment. In addition, many patients experience emotional and psychosocial distress associated with their cancer diagnosis or non-response to therapy. Therefore, symptom control should be an important component of comprehensive and effective therapy.

4.2. Long-Term Lung Cancer Survivors

People who are alive 5 years after a diagnosis of primary lung cancer are referred to as long-term lung cancer (LTLC) survivors *(77)*. Although the chance is only 15%, >25,000 individuals become LTLC survivors every year in the United States *(78)*. Aging of the general population and advancements in early detection and treatment *(79,80)* will further increase the LTLC survivors in the population. Sometimes in the past, over a 30-month survival after an SCLC diagnosis was regarded as LTLC survivors *(81)*. The majority of these LTLC survivors have undergone invasive treatment such as lung resection, radiation therapy, and/or chemotherapy; comorbidity burden in these survivors is especially high compared with survivors of other cancer sites *(82)*. Recurrent disease may occur in a subgroup of LTLC survivors up to >10 years after diagnosis *(10)*, and the survivors are extremely vulnerable (10-fold higher risk than other adult smokers) to developing new aerodigestive tract tumors *(83)*, especially subsequent primary lung cancer (SPLC) and other smoking-related cancers. The Lung Cancer Study Group reported that the incidence of SPLC increased 2-fold after 5 years compared with the preceding 5 years after surgery. The cumulative risk of developing SPLC or other smoking-related cancers reaches 13–20% at 6–8 years *(84)*. Chest radiotherapy and continued smoking were found to significantly increase the risk of SPLC in these patients *(85)*. Late effects of radiation and/or chemotherapy among LTLC survivors have not been defined.

4.2.1. Pulmonary Function Status

Two studies reported the impact of pulmonary function of LTLC survivors. The first study was based on 142 survivors with the observed average FEV_1% predicted being 68% (SD, 23), one-fifth being under 50% predicted FEV_1, and 36% with moderate-to-severe obstructive and/or restrictive ventilatory disorders *(86)*. The second was a 15-year follow-up on the pulmonary status of 156 SCLC patients who had been treated with chest irradiation and chemotherapy that evaluated the time trend of symptoms,

signs, and functions *(87)*. Minimal changes have been found from 2 to 15 years after treatment.

4.2.2. Tobacco and Alcohol Use

Many smokers continue to smoke even after a diagnosis of lung cancer and even after receiving chemotherapy, radiation therapy, or surgery. Thirty percent to 60% of smokers will continue to smoke after their cancer diagnosis *(88)*. In a study of 317 smokers diagnosed with stage I NSCLC, a 2-month tobacco abstinence rate of 53% and a 24-month tobacco abstinence rate of 47% were observed *(89)*. In a pilot study of 148 LTLC survivors, 19% were smoking at the time of diagnosis, but only 5% were still smoking 5 years after diagnosis *(90)*.

Alcohol use was evaluated among lung cancer survivors in a cohort of 142 LTLC survivors *(91)*. At lung cancer diagnosis, 69% consumed alcohol and then 16% reported changes in their alcohol use after diagnosis (either stopped or decreased their amount of alcohol intake). Compared with nondrinkers, an odds ratio of 9.0 showed that drinkers perceived themselves as having poorer health.

4.2.3. Self-Assessed Quality of Life

In a cross-sectional survey of LTLC survivors, fatigue and anxiety were reported as major problems, and their physical functioning scores were worse than other cancer survivors *(92)*. The authors pointed out the importance of the QOL assessment and the pitfalls of assuming QOL findings in the absence of clinical data. Changes in QOL over time have been evaluated among 164 LTLC survivors in a pilot study; 34% of these survivors experienced a significant decline in their overall QOL at the 5-year follow-up compared with their under 3-year follow-up *(93)*.

5. Summary and Perspectives: Knowledge Gap and Research in Need

Outcome research in lung cancer has been focused mainly on short-term survival; there is a shortage of knowledge about the health and quality of life in LTLC survivors at present. Only occasionally in the past, systematic evaluation of survival predictors and QOL attributes were simultaneously conducted in the same study. One such study was carried out in 102 patients with inoperable NSCLC, in which disease symptoms and psychosocial well-being were the best predictors for survival *(94)*. According to the limited information in the literature, the QOL of long-term survivors of lung cancer showed substantial deficits relative to other patient populations, indicating a need for targeted

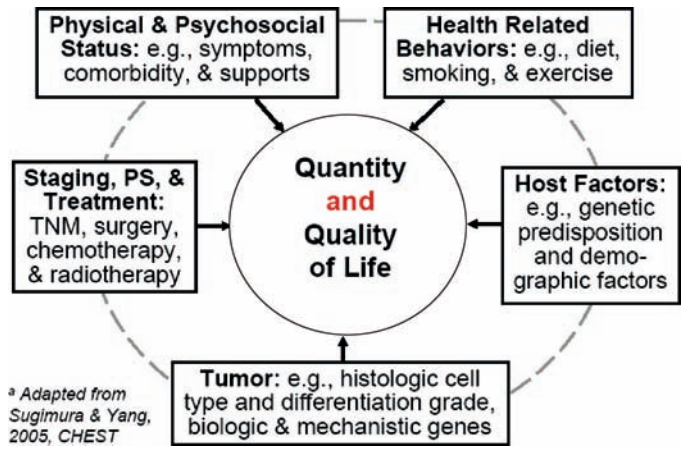


Fig. 24.3. An integrated view: ling cancer outcome research. Adapted from Sugimura and Yang (2005).

interventions *(93)*. QOL has been suggested and should be considered to constitute a prognostic factor for lung cancer survival.

The independent or combined effects of lung cancer treatment, aging, smoking and drinking, comorbid conditions, and psychosocial factors likely cause late effects, including organ malfunction, chronic fatigue, pain, or premature death among LTLC survivors *(95)*. In the mid-1990s, multidimensional models were proposed based on a conceptual framework of Wilson and Cleary *(96)* to capture the most important QOL predictors *(77)*. This framework, adapted to lung cancer survivorship, encompasses the following five dimensions and domains, as illustrated in **Fig. 24.3**:host-related factors (e.g., demographic and genomic), tumor-related factors (e.g., histology and markers of cell proliferation and apoptosis), disease- and treatment-related factors (e.g., adverse effects, symptoms, and disease recurrence), health-related behaviors (e.g., smoking status and physical activity level), co-morbid conditions, and psychosocial facets (e.g., emotional balance and spiritual well-being). With the advanced technology in the genome era, more and more research initiatives are multidisciplinary and overarching basic, clinical, population, and behavioral sciences in achieving the goal of patient-specific medical care. New knowledge gained from these studies could help lung cancer survivors, their healthcare providers, and their caregivers by providing evidence for establishing clinical recommendations to enhance their quantity and quality of life.

References

1. Mountain, C. F. (1997) Revisions in the international system for staging lung cancer. Chest **111**(6), 1710–17.
2. Yang, P., Allen, M. S., Aubry, M. C., Wampfler, J. A., Marks, R. S., Edell, E. S., Thibodeau, S. N., Adjei, A. A., Jett, J., and Deschamps, C. (2005) Clinical features of 5,628 primary lung cancer patients: experience at Mayo Clinic from 1997–2003. Chest **128**, 452–62.
3. Nesbitt, J. C., Lee, J. S., Komaki, R. and Roth, J. A. (1997) Cancer of the lung, in Cancer Medicine (Holland, J. F., Frei, E., III, Bast, R. J., Kufe, D., Morton, D., and Weichselbaum, R., eds.), Williams & Wilkins, Baltimore, MD, pp. 1723–803.
4. Feld, R., Sagman, U., and LeBlanc, M. (1996) Staging and prognostic factors: Small cell lung cancer, in Lung Cancer: Principles and Practice (Pass, H. I., Mitchell, J. B., Johnson, D. H., and Turrisi, A. T., eds.), Lipincott-Raven Publishers, Philadelphia, PA, pp. 495–509.
5. Wozniak, A. J., and Gadgeel, S. M. (2007) Adjuvant treatment of non-small-cell lung cancer: how do we improve the cure rates further? Oncology **21**, 164–71.
6. Rubin, P., Brasacchio, R., and Katz, A. (2006) Solitary metastases: illusion versus reality. Semin. Radiat. Oncol. **16**, 120–30.
7. Martini, N., and Melamed, M. R. (1975) Multiple primary lung cancers. J. Thorac. Cardiovasc. Surg. **70**, 606–12.
8. Scott, W. J. (2004) Metachronous lung cancer: the role of improved postoperative surveillance. J. Thorac. Cardiovasc. Surg. **127**, 633–35.
9. Martini, N., Bains, M. S., Burt, M. E., Zakowski, M. F., McCormack, P., Rusch, V. W., and Ginsberg, R. J. (1995) Incidence of local recurrence and second primary tumors in resected stage I lung cancer. J. Thorac. Cardiovasc. Surg. **109(1)**, 120–29.
10. The Lung Cancer Study Group (1993) Malignant disease appearing late after operation for T1 N0 non-small cell lung cancer. J. Thorac. Cardiovasc. Surg. **106**, 1053–58.
11. Immerman, S. C., Vanecko, R. M., Fry, W. A., Head, L. R., and Shields, T. W. (1981) Site of recurrence in patients with stages I and II carcinoma of the lung resected for cure. Ann. Thorac. Surg. **32**, 23–27.
12. Martini, N., Rusch, V. W., Bains, M. S., Kris, M. G., Downey, R. J., Flehinger, B. J. and Ginsberg, R. J. (1999) Factors influencing ten-year survival in resected stages I to IIIA non-small lung cancer. J. Thorac. Cardiovasc. Surg. **117**, 32–38.
13. Colice, G. L., Rubins, J., and Unger, M. (2003) Follow-up and surveillance of the lung cancer patient following curative-intent therapy. Chest **123**,
14. Thomas, P., Rubinstein, L., and The Lung Cancer Study Group. (1990) Cancer Recurrence after resection: T1 N0 non-small cell lung cancer. Ann. Thorac. Surg. **49**, 242–47.
15. Sugimura, H., Nichols, F. C., Yang, P., Allen, M. S., Deschamps, C., Cassivi, S. D., Williams, B. A., and Pairolero, P. C. (2007) Survival following recurrent non-small cell lung cancer after complete pulmonary resection. Ann Thorac. Surg. **83**, 409–18.
16. Williams, B. A., Sugimura, H., Endo, C., Nichols, F. C., Cassivi, S. D., Allen, M. S., Pairolero, P. C., Deschamps, C., and Yang, P. (2006) Predicting postrecurrence survival among completely resected nonsmall-cell lung cancer patients. Ann.Thorac. Surg. 81, 1021–27.
17. Visbal, A. L., Williams, B. A., Nichols, F. C., Marks, R. S., Jett, J. R., Aubry, M. C., Edell, E. S., Wampfler, J. A., Molina, J. R., and Yang, P. (2004) Gender differences in non-small cell lung cancer survival: an analysis of 4,618 patients diagnosed between 1997–2002. *Ann. Thorac. Surg.* **78(1)**, 209–15.
18. Sun, Z., Aubry, M. C., Deschamps, C., Marks, R. S., Okuno, S. H., Williams, B. A., Sugimura, H., Pankratz, V. S. and Yang, P. (2006) Histologic grade is an independent prognostic factor for survival in non-small cell lung cancer: an analysis of 5018 hospital- and 712 population-based cases. *J. Thorac. Cardiovasc. Surg.* **131**, 1014–20.
19. Ebbert, J. O., Yang, P., Vachon, C. M., Vierkant, R. A., Cerhan, J. R., Folsom, A. R., and Sellers, T. A. (2003) Lung cancer risk reduction after smoking cessation: observations from a prospective cohort of women. *J. Clin. Oncol.* **21(5)**, 921–26.
20. Jatoi, A., Williams, B., Nichols, F. C., Marks, R. S., Aubry, M. C., Wampfler, J., Finke, E. E., and Yang, P. (2005) Is voluntary vitamin and mineral supplementation associated with better outcome in non-small cell lung cancer? results from the Mayo Clinic lung cancer cohort. *Lung Cancer* **49**, 77–84.
21. Jatoi, A., Williams, B. A., Marks, R., Nichols, F. C., Aubry, M. C., Wampfler, J., and Yang, P.

(2005) Exploring vitamin and mineral supplementation and purported clinical effects in patients with small cell lung cancer: results from the Mayo Clinic lung cancer cohort. *Nutr. Cancer* **51**, 7–12.

22. Brundage, M. D., Davies, D., and Mackillop, W. J. (2002) Prognostic factors in non-small cell lung cancer. A decade of progress. *Chest* **122(3)**, 1037–57.
23. Sun, Z., and Yang, P. (2006) Gene expression profiling on lung cancer outcome prediction: present clinical value and future premise. *Cancer Epidemiol. Biomarkers Prev.* **15**, 2063–68.
24. Yanagisawa, K., Shyr, Y., Xu, B. J., Massion, P. P., Larsen, P. H., White, B. C., Roberts, J. R., Edgerton, M., Gonzalez, A., Nadaf, S., Moore, J. H., Caprioli, R. M., and Carbone, D. P. (2003) Proteomic patterns of tumour subsets in non-small-cell lung cancer. *Lancet* **362**, 433–39.
25. Yanagisawa, K., Tomida, S., Shimada, Y., Yatabe, Y., Mitsudomi, T., and Takahashi, T. (2007) A 25-signal proteomic signature and outcome for patients with resected non-small-cell lung cancer. *J. Natl. Cancer Inst.* **99**, 858–67.
26. Saijo, N. (2006) Recent trends in the treatment of advanced lung cancer. *Cancer* **97**, 448–52.
27. Byrne, B. J., and Garst, J. (2005) Epidermal growth factor receptor inhibitors and their role in non-small-cell lung cancer. *Curr. Oncol. Rep.* **7**, 241–47.
28. Shigematsu, H., and Gazdar, A. F. (2006) Somatic mutations of epidermal growth factor receptor signaling pathway in lung cancers. *Int. J. Cancer.* **118**, 257–62.
29. Janne, P. A., Engelman, J. A., and Johnson, B. E. (2005) Epidermal growth factor receptor mutations in non-small-cell lung cancer: implications for treatment and tumor biology. *J. Clin. Oncol.* **23**, 3227–34.
30. Calvo, E., and Baselga, J. (2006) Ethnic differences in response to epidermal growth factor receptor tyrosine kinase inhibitors. *J. Clin. Oncol.* **24**, 2158–63.
31. Jain, R. K. (2003) Molecular regulation of vessel maturation. *Nat. Med.* **9**, 685–93.
32. Molina, J. R., Adjei, A. A., and Jett, J. R. (2006) Advance in chemotherapy of non-small cell lung cancer. *Chest* **130**, 1211–19.
33. Stinchcombe, T. E. and Socinski, M. A. (2007) Bevacizumab in the treatment of non-small-cell lung cancer. *Oncogene* **26**, 3691–98.
34. Ruttinger, D., WInter, H., van den Engel, N. K., Hatz, R. A., Schlemmer, M., Phola, H., Grutzner, S., Schendel, D. J., Fox, B. A., and Jauch, K. W. (2006) Immunotherapy of lung cancer: an update. *Onkologie* **29**, 33–38.
35. Blackhall, F. H., and Shepherd, F. A. (2007) Small cell lung cancer and targeted therapies. *Curr. Opin. Oncol.* **19**, 103–08.
36. Cross, D., and Burmester, J. K. (2006) Gene therapy for cancer treatment: past, present and future. *Clin. Med. Res.* **4**, 218–27.
37. Boulikas, T., and Vougiouka, M. (2003) Cisplatin and platinum drugs at the molecular level. *Oncol. Rep.* **10**, 1163–682.
38. Rogatko, A., Babb, J. S., Wang, H., Slifker, M. J., and Hudes, G. R. (2004) Patient characteristics compete with dose as predictors of acute treatment toxicity in early phase clinical trials. *Clin. Cancer Res.* **10**, 4645–51.
39. Di Maio, M., Gridelli, C., Gallo, C., Shepherd, F., Piantedosi, F., Cigolari, S., Manzione, L., Illiano, A., Barbera, S., Robbiati, S. F., Frontini, L., Piazza, E., Ianniello, G. P., Veltri, E., Castiglione, F., Rosetti, F., Gebbia, V., Seymour, L., Chiodini, P., and Perrone, F. (2005) Chemotherapy-induced neutropenia and treatment efficacy in advanced non-small-cell lung cancer: a pooled analysis of three randomised trials. *Lancet Oncol.* **6**, 669–77.
40. Gurney, H. (2005) I don't underdose my patients . . . do I? *Lancet Oncol.* 6, 637–38.
41. Susman, E. (2004) Rash correlates with tumour respsone after cetuximab. *Lancet Oncol.* **5**, 647.
42. Gurney, H. (2002) How to calculate the dose of chemotherapy. *Br. J. Cancer* **86**, 297–302.
43. Rankin, E. M., Mill, L., Kaye, S. B., Atkinson, R., Cassidy, L., Cordiner, J., Cruickshank, D., Davis, J. R., Duncan, I. D., Fullerton, W., Habeshaw, T., Kennedy, J., Kennedy, R., Kitchener, H., MacLean, A., Paul, J., Reed, N., Sarker, T., Soukop, M., et al. (1992) A randomised study comapring standard dose carboplatin with chlorambucil and carboplatin in advanced ovarian cancer. *Br. J. Cancer* **65**, 275–81.
44. Poikonen, P., Saarto, T., Lundin, J., Joensuu, H., and Blomqvist, C. (1999) Leucocyte nadir as a marker for chemotherapy efficacy in node-positive breast cancer treated with adjuvant CMF. *Br. J. Cancer* **80**, 1763–66.
45. Ratain, M. J., and Relling, M. V. (2001) Gazing into a crystal ball-cancer therapy in

the post-genomic era. *Nat. Med.* **7**, 283–85.

46. Lord, P. G., and Papoian, T. (2004) Genomics and Drug Toxicity. *Science* **306**, 57.
47. Yang, P., Ebbert, J. O., Sun, Z., and Weinshilboum, R. M. (2005) A role of the glutathione metabolic pathway in lung cancer treatment and prognosis: a review. *J. Clin. Oncol.* **24**, 1761–69.
48. Fuertes, M. A., Castilla, J., Alonso, C. and Perez, J. M. (2003) Cisplatin biochemical mechanism of action: from cytotoxicity to induction of cell death through interconnections between apoptotic and necrotic pathways. *Curr. Med. Chem.* **10**, 257–66.
49. Rosell, R., Taron, M., Camps, C., and Lopez-Vivanco, G. (2003) Influence of genetic markers on survival in non-small cell lung cancer. *Drugs Today (Barc)* **39**, 775–86.
50. Zhou, W., Gurubhagavatula, S., Liu, G., Park, S., Neuberg, D. S., Wain, J. C., Lynch, T. J., Su, L., and Christiani, D. C. (2004) Excision repair cross-complementation group 1 polymorphism predicts overall survival in advanced non-small cell lung cancer patients treated with platinum-based chemotherapy. *Clin. Cancer Res.* **10**, 4939–43.
51. Gurubhagavatula, S., Liu, G., Park, S., Zhou, W., Su, L., Wain, J. C., Lynch, T. J., Neuberg, D. S., and Christiani, D. C. (2004) XPD and XRCC1 genetic polymorphisms are prognostic factors in advanced non-small-cell lung cancer patients treated with platinum chemotherapy. *J. Clin. Oncol.* **22**, 2594–601.
52. Ryu, J. S., Hong, Y. C., Han, H. S., Lee, J. E., Kim, S., Park, Y. M., Kim, Y. C., and Hwang, T. S. (2004) Association between polymorphisms of ERCC1 and XPD and survival in non-small-cell lung cancer patients treated with cisplatin combination chemotherapy. *Lung Cancer* **44**, 311–6.
53. Connell, P. P., Kron, S. J., and Weichselbaum, R. R. (2004) Relevance and irrelevance of DNA damage response to radiotherapy. *DNA Repair (Amst)* **3**, 1245–51.
54. Willers, H., Dahm-Daphi, J., and Powell, S. N. (2004) Repair of radiation damage to DNA. *Br. J. Cancer* **90**, 1297–301.
55. Ross, G. M. (2002) Inducation of cell death by radiotherapy. *Endocr. Relat. Cancer* **6**, 41–44.
56. Leslie, E. M., Deeley, R. G., and Cole, S. P. C. (2001) Toxicological relevance of the multidrug resistance protein 1, MRP1 (ABCC1) and related transporters. *Toxicology* **167**, 3–23.
57. Danesi, R., De Braud, F., Fogli, S., Di Paolo, A., and Del Tacca, M. (2001) Pharmacogenetic determinants of anti-cancer drug activity and toxicity. *Trends Pharmacol. Sci.* **22**, 420–26.
58. Strange, R. C., Jones, P. W., and Fryer, A. A. (2000) Glutathione S-trnasferase: genetics and role in toxicology. *Toxicol. Lett.* **112–113**, 357–63.
59. Eaton, D. L., and Bammler, T. K. (1999) Concise Review of the glutathione s-transferases and their significance to toxicology. *Toxicol. Sci.* **49**, 156–64.
60. O'Brien, M. L., and Tew, K. D. (1996) Glutathione and Related Enzymes in Multidrug Resistance. *Eur. J. Cancer* **6**, 967–78.
61. Anonymous. (1995) The World Health Organization Quality of Life Assessment (WHOQOL). Position Paper from the World Health Organization. *Soc. Sci. Med.* **41**, 1403–09.
62. Bonomi, A. E., Patrick, D. L., Bushnell, D. M., and Martin, M. (2000) Validation of the United States' Version of the World Health Organization Quality of Life (WHOQOL) Instrument. *J. Clin. Epidemiol.* **53**, 1–12.
63. Li, W. W. L., Lee, T. W. and Yim, A. P. C. (2004) Quality of life after lung cancer resection. *Thorac. Surg. Clin.* **14**, 353–65.
64. Ware, J. E., and Sherbourne, C. D. (1992) The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med. Care* **30**, 473–83.
65. McHorney, C. A., Ware, J. E., and Raczek, A. E. (1993) The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med. Care* **31**, 247–63.
66. Bergman, B., and Aaronson, N. K. (1995) Quality-of-life and cost-effectiveness assessment in lung cancer. *Curr. Opin. Oncol.* **7**, 138–43.
67. Cella, D. F., Bonomi, A. E., Lloyd, S. R., Tulsky, D. S., Kaplan, E., and Bonomi, P. (1995) Reliability and validity of the Functional Assessment of Cancer Therapy-Lung (FACT-L) quality of life instrument. *Lung Cancer* **12**, 199–200.
68. Hollen, P. J., Gralla, R. J., Kris, M. G., and Cox, C. (1994) Quality of life during clinical trials: conceptual model for the Lung Cancer Symptom Scale (LCSS). *Support. Care Cancer* **2**, 213–22.

69. Montazeri, A., Gillis, C. H., and McEwen, J. (1998) Quality of life in patients with lung cancer. A review of literature from 1970–1995. *Chest* **113(2)**, 467–81.
70. Ganz, P. A., Figlin, R. A., Haskell, C. M., La Soto, N., Siau, J., and The UCLA Solid Tumor Study Group. (1989) Supportive care versus supportive care and combination chemotherapy in metastatic non-small cell lung cancer. *Cancer* **63**, 1271–78.
71. Helsing, M., Bergman, B., Thaning, L., and Hero, U. (1998) Quality of Life and survival in patients with advanced non-small cell lung cancer receiving supportive care plus chemotherapy with carboplatin and etoposide or supportive care only. A multicentre randomized phase III trial. *Eur. J. Cancer.* **34(7)**, 1036–44.
72. Silvestri, G., Pritchard, R., and Weich, H. G. (1998) Preferences for chemotherapy in patients with advanced non-small cell lung cancer: descriptive study based on scripted interviews. *Br. Med. J.* **317**, 771–75.
73. Karmakar, M. K., and Ho, A. M. (2004) Postthoracotomy pain syndrome. *Thorac. Surg. Clin.* **14**, 345–52.
74. Stavraky, K. M., Donner, A. P., Kincade, J. E., and Stewart, M. A. (1988) The effect of psychosocial factors on lung cancer mortality at one year. *J. Clin. Epidemiol.* **41**, 75–82.
75. Reck, M., von Pawel, J., Macha, H. N., Kaukel, E., Deppermann, K. M., Bonnet, R., Ulm, K., Hessler, S., and Gatzemeier, U. (2006) Efficient palliation in patients with small-cell lung cancer by a combination of paclitaxel, etoposide and carboplatin: Quality of life and 6-years'-follow-up results from a randomised phase III trial. *Lung Cancer* **53**, 67–75.
76. Temel, J. S., Pirl, W. F., and Lynch, T. J. (2006) Comprehensive symptom management in patients with advanced-stage non-small-cell lung cancer. *Clin. Lung Cancer* **7**, 241–49.
77. Sugimura, H., and Yang, P. (2006) Long-term survivorship in lung cancer. A review. *Chest* **29**, 1088–97.
78. Institute, N. C. (ed.) (2002) SEER Cancer Statistics Review, 1973–1999. Bethesda, MD. http://seer.cancer.gov/csr/1973-1999/.
79. Winton, T. L., Livingston, R. B., Johnson, D., and et al. (2004) A prospective randomized trial of adjuvant vionrelbine (VIN) and cisplation (CIS) in completely resected stage 1B and II non small cell lung cancer (NSCLC) Intergroup JBR.10. 7018.
80. Strauss, G. M., Herndon, J., Maddaus, M. A., et al. (2004) Randomized clinical trial of adjuvant chemotherapy with paclitzxel and craboplatin following resection in stage IB non-small cell lung cancer (NSCLC): report of Cancer and Leukemia Group B (CALGB) protocol 9633. 7019.
81. Jacoulet, P., Depierre, A., Moro, D., Riviere, A., Milleron, B., Quoix, E., Ranfaing, E., Anthoine, D., Lafitte, J. J., Lebeau, B., Kleisbauer, J. P., F., M., Fournel, P., Zaegel, M., Leclerc, J. P., Garnier, G., Brambilla, E., and Capron, F. (1997) Long-term survivors of small-cell lung cancer (SCLC): a French multicenter study. Groupe d'Oncologie de Langue Francaise. *Ann. Oncol.* **8(10)**, 1009–14.
82. Ko, C. Y., Maggard, M., and Livingston, E. H. (2003) Evaluating health utility in patients with melanoma, breast cancer, colon cancer, and lung cancer: a nationwide, population-based assessment. *J. Surg. Res.* **114**, 1–5.
83. Johnson, B. E., Cortazar, P., and Chute, J. P. (1997) Second lung cancers in patients successfully treated for lung cancer. *Semin. Oncol.* **24**, 492–99.
84. Johnson, B. E. (1998) Second lung cancers in patients after treatment for an initial lung cancer. *J. Natl. Cancer Inst.* **90(18)**, 1335–45.
85. Tucker, M. A., Murray, N., Shaw, E. G., Ettinger, D. S., Mabry, M., Huber, M. H., Feld, R., Shepherd, F. A., Johnson, D. H., Grant, S. C., Aisner, J., and Johnson, B. E. (1997) Second primary cancers related to smoking and treatment of small-cell lung cancer. *J. Natl. Cancer Inst.* **89(23)**, 1782–88.
86. Sarna, L., Evangelista, L., Tashkin, D., Padilla, G., Holmes, C., Brecht, M. L., and Grannis, F. (2004) Impact of respiratory symptoms and pulmonary function on quality of life of long-term survivors of non-small cell lung cancer. *Chest* **125(2)**, 439–45.
87. Myers, J. N., O'Neil, K. M., Walsh, T. E., Hoffmeister, K. J., Venzon, D. J. and Johnson, B. E. (2005) The pulmonary status of patients with limited-stage small cell lung cancer 15 years after treatment with chemotherapy and chest irradiation. *Chest* **128, 3261–68**.
88. Pinto, B. M., Eakin, E., and Maruyama, N. (2000) Health behavior changes after a cancer diagnosis: what do we know and where do we go from here? *Ann. Behav. Med.* **22**, 38–52.
89. Gritz, E. R., Nisenbaum, R., Elashoff, R. E., and Holmes, E. C. (1991) Smoking behavior following diagnosis in patients with stage I

non-small cell lung cancer. *Cancer Causes Control* **2**, 105–12.

90. Yang, P., Sugimura, H., Ebbert, J. O., Nichols, F. C., Kelemen, L. E., Worra, J. B., Stoddard, S. M., Johnson, M. E., and Sloan, J. A. (2004). Characteristics of long-term lung term lung cancer survivors. In: Department of Health and Human Services. National Cancer Institute, and American Cancer Society (eds.), Canon Survirorship: Pathways to Health After Treatment, pp. 51 Washington, DC.
91. Evangelista, L. S., Sarna, L., Brecht, M. L., Padilla, G., and Chen, J. (2003) Health perceptions and risk behaviors of lung cancer survivors. *Heart Lung* **32(2)**, 131–39.
92. Sarna, L., Padilla, G., Holmes, C., Tashkin, D., Brecht, M. L. and Evangelista, L. (2002) Quality of life of long-term survivors of non-small-cell lung cancer. *J. Clin. Oncol.* **20(13)**, 2920–29.
93. Yang, P., Sugimura, H., Sloan, J., Williams, B., Cassivi, S., Garces, Y., Sun, Z., Worra, J., Midthun, D., and Jatoi, A. (2005) Longitudinal evaluation of quality of life in long-term lung cancer survivors (Abstract), in 11th World Conference on Lung Cancer, Presidential Symposium, July 3–6, 2005, Barcelona, Spain. *Lung Cancer* **49**, S4.
94. Kaasa, S., Mastekaasa, A., and Lund, E. (1989) Prognostic factors for patients with inoperable non-small cell lung cancer, limited disease. *Radiother. Oncol.* 15, 235–42.
95. Aziz, N. M. (2002) Cancer survivorship research: challenge and opportunity. *J. Nutr.* **132**, 3494S–503S.
96. Wilson, I. B., and Cleary, P. D. (1995) Linking clinical variables with health-related quality of life: A conceptual model of patient outcomes. *Jama* 27**3,** 59–65.

Chapter 25

Hereditary Breast and Ovarian Cancer Syndrome

The Impact of Race on Uptake of Genetic Counseling and Testing

Michael S. Simon and Nancie Petrucelli

Summary

Breast cancer is a significant cause of morbidity and mortality in the United States. Although breast cancer is more common among White American (WA) women, incidence rates are higher among young African American (AA) women. Approximately 5–10% of all breast cancer can be accounted for by germline mutations in the breast cancer (BRCA)1 and BRCA2 genes responsible for hereditary breast and ovarian cancer (HBOC) syndrome. Although genetic counseling (GC) and genetic testing (GT) for HBOC have become widely accepted by the WA population, cancer genetic services are underused among AA. Many investigators have evaluated a wide spectrum of BRCA1 and BRCA2 mutations in the AA and African population with the possible identification of African founder mutations. Barriers to GC and GT include lack of knowledge and/or negative attitudes regarding genetics and genetics research, and concerns regarding the potential for racial discrimination. It is important for future research to focus on ways in which to eliminate barriers to GC and GT to alleviate disparity in the use of genetic services among high-risk AA women.

Key words: Breast cancer; BRCA1; BRCA2; primary care; survival.

1. Racial Differences in Breast Cancer Incidence, Mortality and Survival

Cancer is a significant cause of morbidity and mortality in the African American (AA) population, with AA men and women sharing a disproportionate share of the cancer burden in the United States *(1,2)*. Breast cancer is the most common cause of cancer among women in the United States, and although breast cancer incidence is higher among White American (WA) than

Mukesh Verma (ed.), *Methods in Molecular Biology, Cancer Epidemiology, vol. 471*
© 2009 Humana Press, a part of Springer Science + Business Media, Totowa, NJ
Book doi: 10.1007/978-1-59745-416-2

AA women, mortality due to breast cancer is higher among AA women *(3)*. Compared with WA women, AA women with breast cancer are diagnosed at a younger age *(4,5)* and have worse prognostic features and lower survival rates *(6–8)*. Racial disparities in breast cancer morbidity and mortality may be in part due to differences in access to care, which is influenced by socioeconomic status as well as cultural and behavioral differences *(9–13)*. Improvement in the availability of screening, early detection, and appropriate medical care for women with breast cancer has the potential to close the racial divide in breast cancer outcomes.

2. Hereditary Breast and Ovarian Cancer Syndrome

It is estimated that cancer risk in 5–10% of women with breast cancer and 10–15% of women with ovarian cancer is associated with germline mutations in the highly penetrant susceptibility genes, breast cancer gene 1 (BRCA1) and BRCA2 *(14,15)*, otherwise known as hereditary breast and ovarian cancer (HBOC) syndrome *(16,17)*. Studies from high-risk families have revealed that women harboring a mutation in either gene have up to an 87% lifetime risk of developing breast cancer, and a 20–44% lifetime risk of ovarian cancer. Breast cancer survivors with inherited mutations in either BRCA1 or BRCA2 are at a substantially increased risk to develop a second breast cancer, as well as ovarian cancer and other types of cancer *(18,19)*. Men who carry a BRCA1 or BRCA2 mutation are at an increased risk to develop breast, prostate, and pancreatic cancer *(18,19)*. **Table 25.1** lists the established "red flags" for HBOC syndrome.

Genetic testing (GT) to identify deleterious BRCA1 and BRCA2 mutations became commercially available in 1996, and it can provide individuals information about cancer risk as well as the opportunity to influence decisions regarding risk reduction strategies, and enable other family members to better define their own risk. Because only 5–15% of breast and ovarian cancer cases are due to hereditary BRCA1 and BRCA2 mutations, GT is not used as a screening test for the general population; thus, the primary question becomes, who needs to be tested. Genetic counseling (GC) for HBOC is widely available in the United States and provides individuals with pre- and posttest counseling, including pedigree interpretation and cancer risk assessment, a discussion regarding the risks, benefits and limitations of GT if medically indicated, and the ordering, interpretation, and disclosure of genetic tests and results.

Table 25.2 lists the recommended management guidelines for BRCA1 and BRCA2 mutation carriers including recommendations

Table 25.1
Red Flags for HBOC

• Breast cancer before age 50
• Ovarian cancer at any age
• Male breast cancer at any age
• Multiple primary cancers
• Ashkenazi Jewish (central/eastern European) ancestry

Table 25.2
Clinical Management of Women with *BRCA* mutations

Increased Surveillance
Breast
• Monthly breast self examinations beginning at age 18 and annual or semiannual clinical breast examinations beginning at age 25. • Yearly mammography and magnetic resonance imaging beginning at age 25.
Ovarian
• Annual or semiannual transvaginal ultrasound and testing for CA-125 to detect ovarian cancer beginning at age 25.
Chemoprevention
Breast
• Drugs such as Tamoxifen significantly reduce the risk of breast cancer in both affected and unaffected mutation carriers.
Ovarian
• Oral contraceptives are associated with up to a 60% risk reduction for ovarian cancer.
Prophylactic Surgery
Breast
• Preventative mastectomy is associated with a > 90% risk reduction for breast cancer.
Ovarian
• Preventative removal of the ovaries and tubes provides approximately a 96% risk reduction for ovarian cancer and up to a 68% reduction in breast cancer risk (if performed in premenopausal women)..

for surveillance, chemoprevention and prophylactic surgery. Domchek et al. *(20)* examined the effect of prophylactic oophorectomy (PO) on mortality among BRCA mutation carriers. They prospectively followed 426 women who were unaffected with either breast or ovarian cancer before their PO who were

age-matched to women who did not undergo PO, and who were unaffected with cancer at the time that their match underwent PO. At a median follow-up of 2–3 years, PO led to a risk reduction for both breast cancer and ovarian cancer consistent with previous reports. In addition, a reduction in breast cancer-specific mortality, ovarian cancer-specific mortality, and overall mortality also was seen. This study illustrates that the early identification of BRCA1 and BRCA2 mutation carriers and institution of early intervention has the potential to reduce cancer rates and save lives.

For AA women, GC and GT for HBOC may be especially important due to higher rates of early onset breast cancer and higher breast cancer mortality rates seen in that population *(3,5)*. Use of GC and GT for HBOC, however, is not common in the AA community, even though several studies have documented the presence of BRCA1 and BRCA2 mutations in high risk AA women *(21,22)*. Results of these studies suggest that the prevalence of BRCA1 and BRCA2 mutations in AA hereditary breast cancer families may be similar to that seen in WA hereditary breast cancer families.

There are no clinical or medical reasons why high-risk AA women should not be referred for GC and GT for HBOC. In a study of 155 high-risk women that underwent GC and GT at the University of Chicago and other centers that participated in the Myriad Genetics beta testing of BRCA1/2 from 1992 to 2003, AA participants were found to have a higher rate of DNA sequence variants than non-AA participants (44.2 vs. 11.5%), but a lower rate of deleterious mutations (27.9 vs. 46.2%). There were no racial differences in the ability of the statistical program (BRCAPRO) to predict the likelihood of a BRCA1 or BRCA2 mutation among AA versus no-AA participants, suggesting that similar clinical criteria can be used to select AA and WA women for GT for HBOC *(23)*.

More widespread use of GC and GT for HBOC among high-risk AA women has the potential to increase early detection, introduce the option of preventive measures, and lower cancer mortality rates.

3. BRCA1/2 Mutation Analyses among Women of African Descent

The majority of GT for BRCA1 and BRCA2 mutations has been conducted among women of WA ancestry *(24)*. Several reports have shown that GC and GT for HBOC syndrome is underused among high-risk AA women *(21,24–26)*. It is thought that the shared genetic background of Africans and U.S.-born AA individuals contributes to a greater susceptibility to early onset breast cancer in both groups *(21)*. BRCA1-associated breast cancers, and breast cancers seen in young AA women share several similar

clinical and pathologic features. Compared with sporadic tumors, BRCA1 associated breast cancers are more likely to be diagnosed at a younger age *(27)*, and they have aggressive features, including poorly differentiated, hormone receptor-negative, medullary histologic type, aneuploid with high S-phase fraction, and high frequency of p-53 mutations *(28–31)*. Breast cancers in young AA women are also more likely to have aggressive clinical and pathologic features *(7,32–37)*.

Many investigators have evaluated high-risk AA and African women for unique mutations in the BRCA1 and BRCA2 genes *(14,21,38–47)*. This was reviewed at the "Summit Meeting on Breast Cancer Among AA Women," where it was reported that of 26 distinct BRCA1 pathogenic mutations identified in AA women, 15 had not been reported previously, and of 18 BRCA2 mutations identified, 10 had not been reported previously *(21)*.

In a study of high-risk AA women at the University of Chicago, five of nine probands had germline BRCA1 mutations, three of which (1832del5, 5296del4, and 3883 insA) were found to be unique *(38)*, and two (1832del5 and 5296del4) also were seen in probands from unrelated families. Other investigators have identified novel BRCA1 mutations, including Met1775Arg found in two unrelated AA families *(14,39)*, and Cys64Gly identified in one AA kindred *(45)*.

In a study of 45 high-risk women diagnosed with breast cancer at Howard University and 92 ethnically matched population based community controls, two protein truncating mutations in BRCA1 were identified (943ins10 and 3450del4) *(40)*. The 943ins10 mutation had been reported previously in a family from the Ivory Coast *(41)*, and in three other families of African ancestry *(42–44)*. The 3450del4 mutation also had been reported in one Norwegian and rwo Canadian families. In another analysis of the BRCA1 gene among 54 AA women with breast cancer unselected for family history (FH) or age, one novel frame-shift mutation (3331insG) was found *(46)*. In a study of 70 high-risk premenopausal women with breast cancer from Nigeria, of which the majority had no known FH of cancer, 4% were found to have deleterious mutations, including two novel BRCA1 mutations (Q1090X and 1742insG) and one protein-truncating mutation in BRCA2 (3034del4) *(47)*.

The results from these studies suggest that BRCA1 and BRCA2 mutations are relatively common among high-risk individuals of African descent, raising the possibility of a link to one or more common African ancestor(s). Another notable feature of these studies is the unique presence of a wide spectrum of sequence variations and mutations in the BRCA1 and BRCA2 genes, which is consistent with a high level of genetic diversity among individuals of African ancestry *(40,46)* . In the Gao et al. *(47)* study, 23% of the participants had sequence variants. These

findings support the presence of BRCA1 and BRCA2 mutations among women of African descent and strengthen the need to improve the availability of GC and GT for HBOC among high-risk AA women.

4. Barriers to Use of BRCA1 and BRCA2 Testing in the AA Population

Underuse of genetic services in the AA community may be in part due to limited general knowledge of genetics *(26,48–55)* and/or negative attitudes regarding genetics and genetics research *(48,49,56,57)*. Other explanations for the underuse of genetics services has to do with concerns that involvement in genetics or GT has the potential to promote racial discrimination *(56–58)*.

Numerous surveys have evaluated knowledge of genetics *(48,49)* and knowledge of GT for HBOC *(26,49–55)*, in the general population and among women identified at high risk for breast cancer comparing responses of AA to WA participants. In a survey of 407 Maryland residents *(48)*, AAs were one half as likely to have ever heard of the mapping of the human genome; and in a survey of 430 adults waiting for jury duty assignment in Philadelphia, AAs were less aware of the availability of predictive GT (25,% of AA vs. 35% of WA respondents having heard of BRCA testing) *(49)*. In another survey of 220 women waiting for routine outpatient medical services at the University of Alabama, AA women knew significantly less about breast cancer, and about their genetic risk for breast cancer, despite adjustment for education and income *(50)*. Finally, a survey designed to assess predictors of awareness and interest in BRCA1 and BRCA2 testing among a population of 400 adult women in a general medical practice in Philadelphia showed that awareness of GT was inversely associated with AA race *(51)*.

Lipkus et al. *(52)* conducted a survey of 266 AA women with and without a FH of breast cancer, and they found overall a poor level of knowledge regarding breast cancer risk factors *(52)*. In a survey of 407 women who had at least one first-degree relative (FDR) with breast and/or ovarian cancer, AA women had significantly lower levels of knowledge of GT than WA women *(53)*. In another survey of 95 adult members of a large AA kindred with a known BRCA1 mutation, overall knowledge regarding HBOC was limited *(26)*. Finally, in a survey among four high-risk groups of women stratified by race and sexual orientation (WA, Lesbian/bisexual, AA, and Ashkenazi Jewish), AA women had the least knowledge of any group regarding GT *(54)*. Only one reported survey of 473 low-risk female HMO members age 50 and older in the Raleigh-Durham-Chapel Hill area showed no differences

in awareness of the discovery of the BRCA1 gene by race *(55)*. The results of these studies suggest that knowledge of genetics and GT for HBOC is lower among AA individuals compared with WA across different segments of the population.

Other studies have focused on attitudes toward genetics and GT for HBOC *(48,49,50,55,56,57,59)*. In an annual general interest survey among 852 adults in the Louisville, KY, metropolitan area, AA respondents were more likely than WA to believe that genetics was harmful for society *(56)* and in the Maryland survey, AA respondents were 3 times as likely as WA respondents to have had a negative reaction to the mapping of the human genome. Specific negative reactions included the idea that genetics researchers were "playing God" or a general distrust in science *(48)*. Among jury duty candidates, AA respondents were more likely to feel that GT would be used by the government to label groups as inferior, and they were less likely to endorse the potential health benefits of GT *(49)*. Among the four groups of high-risk women, AAs had the lowest level of interest in GT, although AA and lesbian women reported beliefs in having more unrestricted access to data on genetic testing *(60)*. Negative attitudes regarding genetics and GT, including issues related to trust in science or trust in the government, also could adversely affect the participation of AA individuals in cancer GC.

A series of nine focus groups evaluating the public's understanding of genetics found that AA respondents were more likely to fear racial discrimination from genetic technology. AA participants in the focus groups were especially concerned that genetic technology could lead to blaming them for diseases that were more prevalent in their community. Both WA and AA were concerned about government and corporate exploitation of genetic technologies *(57)*. Other researchers also have suggested that genetics research could worsen racial disparities in health care by drawing the focus away from issues related to discrimination and poverty, which are believed to be the actual causes of health care disparities *(58)*. Mistrust of genetics in the AA community could also stem from prior experience with the Tuskegee syphilis study *(61)*. Alternatively, AA focus group respondents indicated that participation in genetics research among minority group members could benefit their community by providing opportunities for exposure to state of the art research and technology *(57)*.

Other barriers to GC and GT among high-risk AA women include the lack of knowledge of standard risk factors and a lower perceived personal risk of breast cancer *(50,59)*, while having a high level of concern regarding breast cancer *(50,52,59)*. In the University of Alabama survey, AA participants were especially concerned about discrimination and stigmatization resulting from GT, and the potential lack of confidentiality of test results *(50)*. In the Raleigh-Durham survey, AA women were less interested

than their WA counterparts in GT *(55)*; however, in the Philadelphia general practice survey, there were no racial differences in interest in GT *(51)*. In the Lipkus survey, AA women with a FH of breast cancer were more concerned about their personal risk of breast cancer, and these concerns translated into a higher level of interest in GT *(52)*. In the Fox Chase and Duke survey, AA women reported significantly greater concerns about their personal risk, and they had greater worries about their affected relative *(59)*. This survey also compared scores for "pros" versus "cons" regarding GT, and it found that AA women had higher pros scores, suggesting that they were more likely to state that GT leads to knowledge regarding additional ways in which to prevent cancer. These findings suggest the need to improve education regarding breast cancer risk assessment so that all women have access to an accurate assessment of their personal risks. In addition, educational efforts are needed to effectively communicate that GC and GT for HBOC could not only identify individuals at high risk but also could lead to lower morbidity and mortality for patients and their family members through the use of effective surveillance and prevention measures.

5. Predictors of Use of GC and GT

Several investigators have evaluated factors predictive of the use of genetics services among high-risk AA women *(24,26,62)*. In the study of Kinney et al. *(26)* women without a personal history of breast cancer had a low rate of adherence to breast cancer screening recommendations. However, despite this, 67% of respondents were interested in discussing risk factors for breast cancer, and 82% reported that they would have GT for HBOC if available. Intention to undergo GT was associated with having at least one FDR with breast and/or ovarian cancer, a 50% perceived risk of being a gene carrier, and a lack of knowledge regarding the risk of being a gene carrier. Cost and availability of the test were cited as barriers.

Another investigation evaluated predictors of acceptance of GC and GT for HBOC among 76 high-risk women in Harlem who were offered both GC and GT *(62)*. Women who rejected both GC and GT had significantly less prior knowledge about the genetics of breast cancer than women who accepted both. Women who rejected GC reported greater concerns about stigmatization, and they had higher anticipated levels of negative emotional reactions to positive test results than women who had both. Women who had neither GC or GT demonstrated strong anticipation of guilt among family members. Perceived benefits of GC and GT

among women who tested positive included increased motivation for breast self-examination and increased motivation to help female relatives decide about GT. Barriers to GC included worry about passing the gene to offspring and anxiety about other family members.

Armstrong et al. *(24)* conducted a case control study at the cancer risk evaluation program of the University of Pennsylvania on racial differences in uptake of GT and factors related to referral for GT. AA women with a FH of breast or ovarian cancer were significantly less likely to undergo GC than WA women. These differences were not explained by the predicted probability of carrying a BRCA1 or BRCA2 mutation, socioeconomic status, cancer risk perception and worry, attitudes about the risks and benefits of BRCA1 and BRCA2 testing, or primary care physician discussion.

These studies exemplify the racial divide in use of genetic services. Despite low levels of knowledge and poor compliance with screening, interest in GT seems to exist among high-risk AA women. It is important that proposed interventions to improve use of cancer genetics resources take into account psychological factors as well as the potential impact of GT on the family. Further research is needed to uncover other barriers to use of GC and GT in the AA community.

6. Conclusions

Despite lower breast cancer incidence rates, AA women experience a greater share of morbidity and mortality from breast cancer than WA women in the United States. Numerous studies have documented the presence of both BRCA1 and BRCA2 mutations among individuals of AA and African ancestry. The evidence regarding the prevalence of BRCA1 and 2 mutations as well as the high breast cancer mortality rates experienced in the AA community support the need for making GC and GT more available among high-risk AA women. The literature suggests the presence of a number of barriers to the use of genetic services in the AA community, including lack of knowledge of genetics and GT, adverse attitudes regarding GT and fears of racial discrimination. Educational efforts are needed to improve knowledge regarding genetics to help reduce fears and promote the use of preventative modalities that could potentially save lives.

Improving the comfort levels of health care providers on the topic of genetics and GT will help to promote the use of cancer genetics services among high-risk women. This will become more important in the primary care community as GT becomes widely

available for common adult conditions. A survey of U.S. physicians' attitudes toward GT for cancer found that of 1,251 physicians from eight medical specialties, 84% of oncologists consider themselves qualified to recommend GT, compared with only 40% of primary care doctors *(63)*. Others studies have shown that AA and Latinos use services that requires a doctor's order at a lower rate than do WA. Racial bias and patient preferences contribute to disparities, but their effects seem small. Communication during the medical interaction plays a central role in decision making about subsequent interventions and health behaviors. Research has shown that doctors have poorer communication with minority patients than with others, but problems in doctor-patient communication have received little attention as a potential cause of health disparities *(11)*.

It may take active interventions to decrease disparities in the use of genetic services. A grant at the San Francisco General Hospital provided free GT, educational sessions and multilingual FH questionnaires to influence use of GT among underserved high-risk women. Of 7,316 questionnaires administered, 4,573 were analyzed for ascertainment of high-risk status and 327 were sufficiently high-risk after review. Of those determined to be high risk, 280 (6%) were referred for GC, and 74 underwent GC and GT. The investigators concluded that 40% of patient referrals would have been missed if they had been restricted to the standard mammography registry form previously used in their clinic *(64)*.

In the AA community, it is important to take into account the potential impact of religiosity and spirituality on the uptake of genetics services. In a study of 290 adult breast cancer patients self-referred to the cancer and risk evaluation program at the Lombardi Cancer Center, women that classified themselves as highly spiritual were 80% less likely to receive GT results than less spiritual women. In multivariate analysis, women with a high level of spirituality and low perceived risk of breast cancer were less likely to receive GT results, whereas level of spirituality did not impact women with a high perceived risk of breast cancer *(65)*. It is important to support both the spiritual needs of women at risk as well as to provide information as to how to work with the medical care system to reduce the risk of cancer.

References

1. Merrill, R. M., Weed, D. L. (2001) Measuring the public health burden of cancer in the United States through lifetime and age-conditional risk estimates. *Ann. Epidemiol.* **11,** 547–553.
2. Glanz, K., Croyle, R. T., Chollette, V. Y., Pinn, V. W. (2003) Cancer-related health disparities in women. *Am. J. Public Health* **93,** 292–298.
3. Weir, H. K., Thun, M. J., Hankey, B. F., Ries, L. A., Howe, H. L., Wingo, P. A., Jemal, A., Ward, E., Anderson, R. N., Edwards, B. K. (2003) Annual report to the nation on the status of cancer, 1975–2000, featuring the

uses of surveillance data for cancer prevention and control. *J. Natl. Cancer Inst.* **95,** 1276–1299.

4. Claus, E. B., Risch, N. J., Thompson, W. D. (1990) Age at onset as an indicator of familial risk of breast cancer. *Am. J. Epidemiol.* **131,** 961–972.
5. Simon, M. S., Korczak, J. F., Yee, C. L., Malone, K. E., Ursin, G., Bernstein, L., McDonald, J. A., Deapen, D., Strom, B. L., Press, M. F., Marchbanks, P. A., Burkman, R. T., Weiss, L. K., Schwartz, A. G. (2006) Breast cancer risk estimates for relatives of white and African American women with breast cancer in the Women's Contraceptive and Reproductive Experiences Study. *J. Clin. Oncol.* **24,** 2498–2504.
6. Du, W., Simon, M. S. (2005) Racial disparities in treatment and survival of women with stage I–III breast cancer at a large academic medical center in metropolitan Detroit. *Breast Cancer Res. Treat.* **91,** 243–248.
7. Simon, M. S., Banerjee, M., Crossley-May, H., Vigneau, D., Noone, A. M., Schwartz, K. (2005) Racial differences in breast cancer survival in the Detroit Metropolitan area. *Breast Cancer Res Treat.* **97,** 149–55.
8. Newman, L. A, Griffith, K. A., Jatoi, I., Simon, M. S., Crowe, J. P., Colditz, G. A. (2006) Meta-analysis of survival in African American and white American patients with breast cancer: ethnicity compared with socioeconomic status. *J. Clin. Oncol.* **24,** 1342–1349.
9. Shavers, V. L., Brown, M. L. (2002) Racial and ethnic disparities in the receipt of cancer treatment. *J. Natl. Cancer Inst.* **94,** 334–357.
10. O'Malley, C. D., Le, G. M., Glaser, S. L., Shema, S. J., West, D. W. (2003) Socioeconomic status and breast carcinoma survival in four racial/ethnic groups: a population-based study. *Cancer* **97,** 1303–1311.
11. Ashton, C. M, Haidet, P., Paterniti, D. A., Collins, T. C., Gordon, H. S., O'Malley, K., Petersen, L. A., Sharf, B. F., Suarez-Almazor, M. E., Wray, N. P., Street, R. L Jr.,. (2003) Racial and ethnic disparities in the use of health services: bias, preferences, or poor communication? *J. Gen. Intern. Med.* **18,** 146–152.
12. Joslyn, S. A. (2002) Racial differences in treatment and survival from early-stage breast carcinoma. *Cancer* **95,** 1759–1766.
13. Shavers, V. L., Harlan, L. C., Stevens, J. L. (2003) Racial/ethnic variation in clinical presentation, treatment, and survival among breast cancer patients under age 35. *Cancer* **97,** 134–147.
14. Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L. M., Ding, W. (1994) A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **266,** 66–71.
15. Wooster, R., Bignell, G., Lancaster, J., Swift, S., Seal, S., Mangion, J., Collins, N., Gregory, S., Gumbs, C., Micklem, G. (1995) Identification of the breast cancer susceptibility gene BRCA2. *Nature* **378,** 789–792.
16. Rebbeck, T. R., Lynch, H. T., Neuhausen, S. L., Narod, S. A., Van't Veer, L., Garber, J. E., Evans, G., Isaacs, C., Daly, M. B., Matloff, E., Olopade, O. I., Weber, B. L. (2002) Prophylactic oophorectomy in carriers of BRCA1 or BRCA2 mutations. *N. Engl. J. Med.* **346,** 1616–1622.
17. Rebbeck, T. R, Friebel, T., Lynch, H. T., Neuhausen, S. L., van't Veer L, Garber, J. E., Evans, G. R., Narod, S. A., Isaacs, C., Matloff, E., Daly, M. B., Olopade, O. I., Weber, B. L. (2004) Bilateral prophylactic mastectomy reduces breast cancer risk in BRCA1 and BRCA2 mutation carriers: the PROSE Study Group. *J. Clin. Oncol.* **22,** 1055–1062.
18. Ford, D., Easton, D. F., Bishop, D. T., Narod, S. A., Goldgar, D. E. (1994) Risks of cancer in BRCA1-mutation carriers. Breast Cancer Linkage Consortium. *Lancet* **343,** 692–695.
19. Ford, D., Easton, D. F., Stratton, M., Narod, S., Goldgar, D., Devilee, P., Bishop, D. T., Weber, B., Lenoir, G., Chang-Claude, J., Sobol, H., Teare, M. D., Struewing, J., Arason, A., Scherneck, S., Peto, J., Rebbeck, T. R., Tonin, P., Neuhausen, S., Barkardottir, R., Eyfjord, J., Lynch, H., Ponder, B. A., Gayther, S. A., Zelada-Hedman, M. (1998) Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium. *Am. J. Hum. Genet.* **62,** 676–689.
20. Domchek, S. M., Friebel, T. M., Neuhausen, S. L., Wagner, T., Evans, G., Isaacs, C., Garber, J. E., Daly, M. B., Eeles, R., Matloff, E., Tomlinson, G. E., Van't, V. L., Lynch, H. T., Olopade, O. I., Weber, B. L., Rebbeck, T. R. (2006) Mortality after bilateral salpingo-oophorectomy in BRCA1 and BRCA2 mutation carriers: a prospective cohort study. *Lancet Oncol.* **7,** 223–229.
21. Olopade, O. I., Fackenthal, J. D., Dunston, G., Tainsky, M. A., Collins, F., Whitfield-Broome, C. (2003) Breast cancer genetics in African Americans. *Cancer* 97, 236–245.

22. Ademuyiwa, F. O., Olopade, O. I. (2003) Racial differences in genetic factors associated with breast cancer. *Cancer Metastasis Rev.* **22,** 47–53.
23. Nanda, R., Schumm, L. P., Cummings, S., Fackenthal, J. D., Sveen, L., Ademuyiwa, F., Cobleigh, M., Esserman, L., Lindor, N. M., Neuhausen, S. L., Olopade, O. I. (2005) Genetic testing in an ethnically diverse cohort of high-risk women: a comparative analysis of BRCA1 and BRCA2 mutations in American families of European and African ancestry. *Jama* **294,** 1925–1933.
24. Armstrong, K., Micco, E., Carney, A., Stopfer, J., Putt, M. (2005) Racial differences in the use of BRCA1/2 testing among women with a family history of breast or ovarian cancer. *Jama* **293,** 1729–1736.
25. Matthews, A. K., Cummings, S., Thompson, S., Wohl, V., List, M., Olopade, O. I. (2000) Genetic testing of African Americans for susceptibility to inherited cancers: use of focus groups to determine factors contributing to participation. *J. Psychosoc. Oncol.* 18, 1–19.
26. Kinney, A. Y., Croyle, R. T., Dudley, W. N., Bailey, C. A., Pelias, M. K., Neuhausen, S. L. (2001) Knowledge, attitudes, and interest in breast-ovarian cancer gene testing: a survey of a large African-American kindred with a BRCA1 mutation. *Prev. Med.* **33,** 543–551.
27. Couch, F. J., DeShano, M. L., Blackwood, M. A., Calzone, K., Stopfer, J., Campeau, L., Ganguly, A., Rebbeck, T., Weber, B. L. (1997) BRCA1 mutations in women attending clinics that evaluate the risk of breast cancer. *N. Engl. J. Med.* 336, 1409–1415.
28. Phillips, K. A., Nichol, K., Ozcelik, H., Knight, J., Done, S. J., Goodwin, P. J., Andrulis, I. L. (1999) Frequency of p53 mutations in breast carcinomas from Ashkenazi Jewish carriers of BRCA1 mutations. *J. Natl. Cancer Inst.* **91,** 469–473.
29. Marcus, J. N., Watson, P., Page, D. L., Narod, S. A., Tonin, P., Lenoir, G. M., Serova, O., Lynch, H. T. (1997) BRCA2 hereditary breast cancer pathophenotype. *Breast Cancer Res. Treat.* **44,** 275–277.
30. (1997) Pathology of familial breast cancer: differences between breast cancers in carriers of BRCA1 or BRCA2 mutations and sporadic cases. Breast Cancer Linkage Consortium. *Lancet* **349,** 1505–1510.
31. Verhoog, L. C., Brekelmans, C. T., Seynaeve, C., van den Bosch, L. M., Dahmen, G., van Geel, A. N., Tilanus-Linthorst, M. M., Bartels, C. C., Wagner, A., van den, O. A., Devilee, P., Meijers-Heijboer, E. J., Klijn, J. G. (1998) Survival and tumour characteristics of breast-cancer patients with germline mutations of BRCA1. *Lancet* **351,** 316–321.
32. Chen, V. W., Correa, P., Kurman, R. J., Wu, X. C., Eley, J. W., Austin, D., Muss, H., Hunter, C. P., Redmond, C., Sobhan, M.,. (1994) Histological characteristics of breast carcinoma in blacks and whites. *Cancer Epidemiol.Biomarkers Prev.* **3,** 127–135.
33. Joslyn, S. A. (2002) Hormone receptors in breast cancer: racial differences in distribution and survival. *Breast Cancer Res. Treat.* 73, 45–59.
34. Simon, M. S., Severson, R. K. (1996) Racial differences in survival of female breast cancer in the Detroit metropolitan area. *Cancer* **77,** 308–314.
35. Jones, B. A., Kasl, S. V., Howe, C. L., Lachman, M., Dubrow, R., Curnen, M. M., Soler-Vila, H., Beeghly, A., Duan, F., Owens, P. (2004) African-American/White differences in breast carcinoma: p53 alterations and other tumor characteristics. *Cancer* **101,** 1293–1301.
36. Li, C. I., Malone, K. E., Daling, J. R. (2003) Differences in breast cancer stage, treatment, and survival by race and ethnicity. *Arch. Intern. Med.* **163,** 49–56.
37. Grann, V., Troxel, A. B., Zojwalla, N., Hershman, D., Glied, S. A., Jacobson, J. S. (2005) Regional and racial disparities in breast cancer-specific mortality. *Soc. Sci. Med.*
38. Gao, Q., Neuhausen, S., Cummings, S., Luce, M., Olopade, O. I. (1997) Recurrent germ-line BRCA1 mutations in extended African American families with early-onset breast cancer. *Am. J. Hum. Genet.* **60,** 1233–1236.
39. Futreal, P. A., Liu, Q., Shattuck-Eidens, D., Cochran, C., Harshman, K., Tavtigian, S., Bennett, L. M., Haugen-Strano, A., Swensen, J., Miki, Y. (1994) BRCA1 mutations in primary breast and ovarian carcinomas. *Science* **266,** 120–122.
40. Panguluri, R. C., Brody, L. C., Modali, R., Utley, K., Adams-Campbell, L., Day, A. A., Whitfield-Broome, C., Dunston, G. M. (1999) BRCA1 mutations in African Americans. *Hum. Genet.* **105,** 28–31.
41. Stoppa-Lyonnet, D., Laurent-Puig, P., Essioux, L., Pages, S., Ithier, G., Ligot, L., Fourquet, A., Salmon, R. J., Clough, K. B., Pouillart, P., Bonaiti-Pellie, C., Thomas, G. (1997) BRCA1 sequence variations in 160 individuals referred to a breast/ovarian family cancer clinic. Institut Curie Breast

Cancer Group. *Am. J. Hum. Genet.* **60,** 1021–1030.

42. Arena, J. F., Smith, S., Plewinska, M. (1996) BRCA1 mutations in African American women. *Am. J. Hum. Genet.* **59(Suppl)** A34.
43. Arena, J. F., Smith, S., Vincek, V. (1997) A BRCA1 founder mutation in African Americans. *Am. J. Hum. Genet.* **65(Suppl)** A325.
44. Mefford, H. C., Baumbach, L., Panguluri, R. C., Whitfield-Broome, C., Szabo, C., Smith, S., King, M. C., Dunston, G., Stoppa-Lyonnet, D., Arena, F. (1999) Evidence for a BRCA1 founder mutation in families of West African ancestry. *Am. J. Hum. Genet.* **65,** 575–578.
45. Castilla, L. H., Couch, F. J., Erdos, M. R., Hoskins, K. F., Calzone, K., Garber, J. E., Boyd, J., Lubin, M. B., DeShano, M. L., Brody, L. C. (1994) Mutations in the BRCA1 gene in families with early-onset breast and ovarian cancer. *Nat. Genet.* 8, 387–391.
46. Shen, D., Wu, Y., Subbarao, M., Bhat, H., Chillar, R., Vadgama, J. V. (2000) Mutation analysis of BRCA1 gene in African-American patients with breast cancer. *J. Natl. Med. Assoc.* **92,** 29–35.
47. Gao, Q., Adebamowo, C. A., Fackenthal, J., Das, S., Sveen, L., Falusi, A. G., Olopade, O. I. (2000) Protein truncating BRCA1 and BRCA2 mutations in African women with pre-menopausal breast cancer. *Hum. Genet.* **107,** 192–194
48. Tambor, E. S., Bernhardt, B. A., Rodgers, J., Holtzman, N. A., Geller, G. (2002) Mapping the human genome: an assessment of media coverage and public reaction. *Genet. Med.* **4,** 31–36.
49. Peters, N., Rose, A., Armstrong, K. (2004) The association between race and attitudes about predictive genetic testing. *Cancer Epidemiol. Biomarkers Prev.* **13,** 361–365.
50. Donovan, K. A., Tucker, D. C. (2000) Knowledge about genetic risk for breast cancer and perceptions of genetic testing in a sociodemographically diverse sample. *J. Behav. Med.* **23,** 15–36.
51. Armstrong, K., Weber, B., Ubel, P. A., Guerra, C., Schwartz, J. S. (2002) Interest in BRCA1/2 testing in a primary care population. *Prev. Med.* **34,** 590–595.
52. Lipkus, I. M., Iden, D., Terrenoire, J., Feaganes, J. R. (1999) Relationships among breast cancer concern, risk perceptions, and interest in genetic testing for breast cancer susceptibility among African-American women with and without a family history of breast cancer. *Cancer Epidemiol. Biomarkers Prev.* **8,** 533–539.
53. Hughes, C., Gomez-Caminero, A., Benkendorf, J., Kerner, J., Isaacs, C., Barter, J., Lerman, C. (1997) Ethnic differences in knowledge and attitudes about BRCA1 testing in women at increased risk. *Patient Educ. Couns.* **32,** 51–62.
54. Durfy, S. J., Bowen, D. J., McTiernan, A., Sporleder, J., Burke, W. (1999) Attitudes and interest in genetic testing for breast and ovarian cancer susceptibility in diverse groups of women in western Washington. *Cancer Epidemiol. Biomarkers Prev.* **8,** 369–375.
55. Tambor, E. S., Rimer, B. K., Strigo, T. S. (1997) Genetic testing for breast cancer susceptibility: awareness and interest among women in the general population. *Am. J. Med. Genet.* **68,** 43–49.
56. Furr, L. A. (2002) Perceptions of genetics research as harmful to society: differences among samples of African-Americans and European-Americans. *Genet. Test.* **6,** 25–30.
57. Bates, B. R., Lynch, J. A., Bevan, J. L., Condit, C. M. (2005) Warranted concerns, warranted outlooks: a focus group study of public understandings of genetic research. *Soc. Sci. Med.* **60,** 331–344.
58. Sankar, P., Cho, M. K., Condit, C. M., Hunt, L. M., Koenig, B., Marshall, P., Lee, S. S., Spicer, P. (2004) Genetic research and health disparities. *Jama* **291,** 2985–2989.
59. Hughes, C., Lerman, C., Lustbader, E. (1996) Ethnic differences in risk perception among women at increased risk for breast cancer. *Breast Cancer Res Treat.* **40,** 25–35.
60. Durfy, S. J., Bowen, D. J., McTiernan, A., Sporleder, J., Burke, W. (1999) Attitudes and interest in genetic testing for breast and ovarian cancer susceptibility in diverse groups of women in western Washington. *Cancer Epidemiol. Biomarkers Prev.* **8,** 369–375.
61. Brandon, D. T., Isaac, L. A., LaVeist, T. A. (2005) The legacy of Tuskegee and trust in medical care: is Tuskegee responsible for race differences in mistrust of medical care? *J. Natl. Med. Assoc.* **97,** 951–956.
62. Thompson, H. S., Valdimarsdottir, H. B., Duteau-Buck, C., Guevarra, J., Bovbjerg, D. H., Richmond-Avellaneda, C., Amarel, D., Godfrey, D., Brown, K., Offit, K. (2002) Psychosocial predictors of BRCA counseling and testing decisions among urban African-American women. *Cancer Epidemiol. Biomarkers Prev.* **11,** 1579–1585.
63. Freedman, A. N., Wideroff, L., Olson, L., Davis, W., Klabunde, C., Srinath, K. P., Reeve, B. B., Croyle, R. T., Ballard-Barbash, R.

(2003) US physicians' attitudes toward genetic testing for cancer susceptibility. *Am. J. Med. Genet.* **120,** 63–71.

64. Lee, R., Beattie, M., Crawford, B., Mak, J., Stewart, N., Komaromy, M., Esserman, L., Shaw, L., McLennan, J., Strachowski, L., Luce, J., Ziegler, J. (2005) Recruitment, genetic counseling, and BRCA testing for underserved women at a public hospital. *Genet. Test.* **9,** 306–312.
65. Schwartz, M. D., Hughes, C., Roth, J., Main, D., Peshkin, B. N., Isaacs, C., Kavanagh, C., Lerman, C. (2000) Spiritual faith and genetic testing decisions among high-risk breast cancer probands. *Cancer Epidemiol. Biomarkers Prev.* **9,** 381–385.

INDEX

A

Absolute excess risk (AER) 89
Acetylation 281–283, 463
Acquired immunodeficiency syndrome (AIDS) 388, 410
Adducts..... 306
Adenomatous polyposis coli (APC) 163–164
Adenosarcoma 142–143
Adenosine diphosphate ribosyl transferase (ADPRT) 305–307, 309, 312, 326
Adjuvant therapy 55–56
Admixed populations..... 232
African..... 387–388
Age-adjusted incidence rate 35
Age-specific incidence..... 51
AIDS associated malignancy..... 387–388
Alaska Native 67–68
Alcohol
  intake..... 92–93, 165, 338–339, 409–410, 423, 480
  related second malignancies..... 92–93
Alkylating agent 96
American Cancer Society..... 439–440
American Caucacians 312, 318
American Indians 67–68
American Joint Committee on cancer (AJCC) 205
Angiogenesis 387–388
  angiogenic feature..... 388–389
Antibody response 394
Asbestos exposure..... 164–165
Ascertainment 416–418
Asians..... 68, 71, 74, 77, 229, 231–233
Astrocytoma tissues 459
Asymptomatic people..... 108–109
Asymptomatic viral carrier 390–391
Ataxia telangiectasia mutated (ATM) 378
ATP production 291–292
Attributable risk 169–170, 240
Attribution bias 25

B

Base excision repair (BER)..... 362–374
Behavioral differences..... 487–488
Behavioral risk factors 90
  surveillance system..... 33
Benzo(a)pyrene (BAP) 370, 372
Binary indicator..... 179–180
Biofluids..... 198–199
Bioinformatics 99, 199
Biomarker..... 99, 165–167
Biospecimen collection 99
BK Virus (BKV) 388
Blacks (African Americans)..... 55–56, 67–68
Bladder cancer 96
Bleomycin sensitivity 372
Bloom and Richardson grading system 56
Body mass index (BMI) 181, 336
Bone sarcoma 96
Bovine leukemia virus..... 433
Brain tumor 90–91
BRCA1 and BRCA2..... 163–164, 487–488, 492
Breast cancer..... 14, 71, 139, 312, 318
  breast self examination..... 58–59
  breast self-awareness..... 58–59
  mortality 58–59
Breast Health Global Initiative (BHGI) 60
British Doctors study..... 218–219

C

Cancer
  ascertainment..... 417–418
  burden..... 5
  death rate..... 70–71
  diagnosis 85–86
  etiology 32
  incidence rate..... 495
  information system 57–58
  inherited syndromes 90–92
  registries..... 4–5, 86–89, 118, 127
  related health outcome..... 65
  specific susceptibility 90–92
  survival..... 57–58, 71–76
  susceptibility genes 305–306
  therapy..... 90
Carbonylation..... 280–281
Carcinogen 137–138
Carcinogenesis..... 305–306
Case-control study..... 90, 172, 263–266
Case-only study..... 172–173
Caucasian 55–57
Cause specific approach..... 253–254
Cell cycle regulatory protein..... 443–444

Center for Medicare and Medicaid Services (CMS) 34–35
Centrosomal abnormalities 443
Cervical cancer 18, 20–21, 439–440
Cervical intraepithelial neoplasia (CIN) 440
Cervix uteri 94–95
Chemotherapy 95–96
Chromatid remodeling 463–464
Chromatin
  immunoprecipitation 287
  modeling 273–274
Chromosome
  chromosomal abnormalities 457–458
  chromosomal duplication 443
Chronic inflammation 175–176
Circulating estrogen 54
Citrullination 280–281
Citrus fruits 93
City registry 264
Classification and regression tree (CART) analysis 174
Closed cohort 219
Cohort Studies 218–221
Colon cancer 12, 17–18, 71–72
Colposcopy 449–450
Comet assay 165–166
Computed tomography (CT) 111
Conditional logistic regression 249–251, 268–269
Condylomatas 440
Contraception 94
Control arm 479
Controlled trial 58–59
Cox model 254–258, 266
Cox regression 180, 187–189
CpG Island (CpG) 286, 458
Cultural and logistic barriers 59
Cumulative incidence 89
Cyclin D1 342–343
Cyclo-oxygenase-2 (COX-2) 337–341, 343, 345–351
Cysteinylglycine 201–202
Cytomegalovirus 388–389

D

Data collection 68
Demethylase 463
Demographic factors 13–24
Department of Health and Human Services (DHHS) 234–235
Detection 199
Determinants 417
Detoxification 165–166, 476
Developed countries 58–59
Diagnosis 6–7, 107–109
Diet 71
  dietary factor 76–77
  dietary supplements 472
Direct reversal 362, 377–378
Disease
  control 70
  factors 58
  progression 107–109
Disparity 69–70
Displacement loop (D-loop) 297–299, 301
Diverse population 69–70, 76–79
Dizygotic twin 165
DNA
  adducts 165–166
  base excision repair 306
  hypermethylation 457–463
  microarray 473
  repair 306, 362, 476
  repair pathway 457–458
  repair protein 312
Double strand break repair (DSBR) 362, 375–377
Drinking 15, 17
Ductal carcinoma 55

E

Early detection 57–59, 97–98, 336–337, 487–488
Early diagnosis 107–109
Eastern European 387–388
Educational gradient 15–16
Elderly 75–76, 78–79
Encode 286
Endometrial cancer 97
Endoscopy 78, 142
Environment
  environmental factors 3, 17, 151, 153, 155, 163–176, 224
  pollution 337–338
Epidemiology and public health 68, 76–79
Epidermal growth factor (EGF) 202–203
Epidermal growth factor receptor (EGFR) 337–344
Epigenetics 433
  epigenetic biomarker 165–167
  epigenetic changes 457–458, 464–465
  epigenetic silencing 457–458
Episome 399
Epstein Barr Virus (EBV) 95, 433
Esophageal cancer 92, 335–352
Estrogen 423
Estrogen receptor 55–56, 424–425
Ethnicity 3, 17–18, 65–66, 165, 227–235
  ethnic groups 53
Etiology 32, 217, 423–424
  etiologic factors 92

Exposure 137–138
field 96
frequencies 218

F

False ER negative results 55–56
False positive 59
Family history 165
Fecal occult blood test 78, 111
Fetal life 423–424

G

Gallbladder cancer 14, 199, 201
Gamma herpesvirus subfamily 390
Gastroesophageal reflex disease (GERD) 336, 339–340
Generalized Estimating Equation 260
Genes
environment interaction 163, 167–174
genetic alterations 457–458
genetic factors 90, 143, 165
genetic polymorphism 163
genetic predisposition 475–477
genetic susceptibility 17–18
penetrance 163
Genistein 175–176
Genomic patterns 474–475
Genomic variants 396–397
Genotypic heterogeneity 66–67
Geographic information system (GIS) 21–22, 40–41, 146–147
Geographic location 53, 66
Geographic variability 388–389
Geographic variation 143–148
Germline mutation 90–92, 488
Global hypomethylation 457–458
Glucocorticoid response element 427
Glucocorticoids 426–427
Glutamyltranspeptidase 201–202
Glutathione 201–202
Glycoproteomics 199
Glycosylation 280–281
Grading system 56
Green tea 175–176
Growth factor receptor 336–337

H

Haplogroup 292–293
Haplotype 292–295
analysis 326–327
Hazard function 188
Hazards ratio 240
HCFA common procedure coding system 34–35
Head and neck cancer 86–87
Health
damaging behaviors 69–70
disparity 69–70, 146, 234
insurance 15, 17
related risk behaviour 78–79
Health and human services (HHS) 69
Healthy people 79–80
Helicobacter pylori 165
Hepatectomy 184–185
Her-2-neu 56–57
Heritable cancer susceptibility 90–91
Herpesviral capsid 389
High risk families 487
High throughput 198
Highly-active Antiretroviral Therapy (HAART) 387–388, 412
Hispanics 68
Histone
DNA interaction 280–281
methylation 463
modification 165–166, 278–284, 463–464
proteins 463
Histone acetyl transferase 280–281
Histone deacetylases (HDACS) 463
Histopathology 55–57
HIV 409–413
Hodgkin's disease 95
Hormone
hormonal factor 90, 94
receptor 55–57
therapy 55–56, 95–97
Hormone replacement therapy (HRT) 22–23, 94
Hormone responsive elements (HRE) 428
Hospital-based cancer registry 138–139
Host cell reactivation assay 165–166
Host factors 58, 90
Host susceptibility 18, 20
Human 8-oxoguanine DNA glycosylase 363–372
Human Herpesvirus-8 (HHV8) 387–400
Human Papilloma Virus (HPV) 94–95, 424, 426–428, 439–452
Hypermethylation 175–176, 273–274
Hypertension 474–475
Hypopharyngeal cancer survivors 92–93

I

Icosahedral capsid 441–443
IGF binding protein 3 199
Immunoblotting 205
Immunocompetent individuals 387–388
Immunodeficiency 376
Immunodepressed population 410

Immunosuppression 417, 439–440
immunosuppressed patients 387–388
immunosuppressive therapy 387–388
Immunotherapy 475
In situ cervical cancer 94–95
Incidence 4–5, 35–37, 56, 137–138, 469–470
rate 35–37
Independent variable 180
Index tumor 92
Indians 466–467
Infection-related cancer 95
Infiltrating ductal carcinoma 55
Inflammation 387–388
Infusion centers 60
Institute of Medicine 67–68
Insulin-like growth factor (IGF-1) 199
Interaction contrast ratio (ICR) 167–168
International classification of diseases (ICD) codes 34–35
International classification of diseases and health related problems-10th edition (ICD-10) 412–413
Intraepithelial neoplasia 440
Intrafamilial transmission 393
Intron 292
Ionizing radiations 476–477
Isoforms 363, 371–372
Isotope-coded affinity tags (ICATs) 207–208
Italian HIV Seroconversion Study (ISS) 410

J

Jewish heritage 387
Joint effect 169

K

Kaposi's sarcoma 91, 95–96, 387–401, 409
Kaposi's sarcoma-associated Herpesvirus (KHSV) 387–401
Keratinocytes 441, 443

L

Lab-on-a-chip 285
Laser capture microdissection (LCM) 205, 207
Latent cycle 390–391
Latinos 67, 70, 72
Lead-time bias 11
Leukemia 12–14, 24, 91, 96, 98–99
Life expectancy 57
Life style 85, 90, 95
factors 143, 151, 153, 164, 410, 416
Linear and Poisson Regression Models 36
Liver cancer 14, 17–18, 24
Lobular carcinoma 55
Logistic regression 179–180, 185–187
Long and short interspersed repeats (LINE and SINE) 463
Long Control Region (LCR) 443
Lung cancer 9, 12, 14, 20, 22–24, 27, 65–66, 71, 74, 78, 469–581
Lymphadenopathic disease 388
Lymphocytic infiltrates 391
Lymphoma 387, 389, 399
Lymphotrophic virus 388–389
Lynch Syndrome (Hereditary Nonpolyposis Colorectal Cancer, HNPCC) 92
Lytic cycle 390

M

Mammography 23, 58, 108–109, 111, 123–127
Mass spectrometry (MS) 197, 208
Matrix assisted desorption/ionization time of flight mass spectrometry (MALDI TOF MS) 473
MeCP2 463–464
Medicade 228
Medical care 488
Medical records 89–90
Medically underserved 68–69
Medicare 31, 33–35, 228, 232, 235
Mediterranian 387, 393–394
Menopause 53–54, 94, 425–426
Metachronous 86
Metastatic status 473, 478
Methylated DNA immunoprecipitation pattern (MeDIP) 286
Methylating agents 306
Methylation modifications 166
Methylation pattern 273, 277, 280
Methylguanine DNA methyl transferase (MGMT) 377–378, 459–462
Micro RNAs (miRNAs) 399
Microfluidics 208
Migrant studies 17–18, 53, 151–153
Minority and underserved population 69
Mismatch repair (MMR) 362
Mitochondria 291–301
Modality 58
Monocytes 391
Monozygotic twin 165
Morbidity 140–141, 145
Mortality 3–5, 35–36, 71, 139, 469–470
rate 58–59, 71
registration 57
Mouse Mammary Tumor Virus (MMTV) 424, 426–434
Multifactor dimensionality reduction (MDR) 174
Multifactorial regression 180
Multifocal neoplasm 387–388

Multiparity 440
Multiple cancers 85–86
Multiple malignancies 97–98
Multiple pigmented sarcoma 387–388
Multiplexing of markers 198
Multivariate analysis 224–225
Multivariate regression 171
Mutagen sensitivity 166
Mutation 90–92, 293–300
Mycoplasma 388–389
Myeloma 227–228

N

National Center for Health Statistics (NCHS) 33, 41
National Health and Nutrition Examination Survey (NHANES) 448–449
National Institutes of Health (NIH) 67–68
National Program of Cancer Registries (NPCR) 33–34
Native American 232
Native Hawaiian 68
Nested case control design 223
Neuroblastoma 13–14
Nodal status 56
Non-coding region (NCR) 443
Non-hispanic Whites 55–56
Non-Hodgkin lymphoma 95, 409–410
Non-melanoma skin cancer 409–410, 413
Non-small cell lung carcinoma 200, 202–203
North-American Association of Central Cancer Registries (NAACCR) 232–233
Nucleosome-nucleosome interaction 281
Nucleotide excision repair (NER) 362
Nulliparous women 447
Nurse's Health Study I and II 201–202

O

Obesity 93–94, 175, 339
Observational studies 76, 112, 114, 119–120
Oncogene 175–176, 399
  oncogenic viral etiology 423–424
  oncogenic viruses 95
Open cohort 219
Open reading frame (ORF) 396
Oral cavity 92–93
Organ transplantation 95, 388, 392–393
Organochlorine 424
Oropharynx 94–95, 392
Outcome variable 180
Ovarian cancer 91
Overdiagnosis 108, 120–124, 126–127, 129–131
Oxidative stress 176
Oxydative Phosphorylation (OXYPHOS) 292
  OXOPHOS complex 292

P

Pacific Islanders 67–68
Pancreatic cancer 7, 11
Papanicolaou 450
Papilloma virus 388–389
Parity 94
Patient entitlement and diagnosis summary file (PEDSF) 34–35
Penetrance 90–91
Peptide mass fingerprint 198
Pharmacogenomic studies 475–476
Pharynx 92
Phosphorylation 280–281
Physical activity 93
Ploidy 472
Poly-ADP ribosylation 280–281, 307
Polygenic model 92
Polymorphism 362
Population
  based interventions 76–79
  based mammography screening 58
  pyramid 53
Positive predictive value 59
Positron emission tomography (PET) 111
Post-translational modifications 284
Power 241–242
Power estimation 251
Pre-disease status 107–108
Prediction 199
Predictor variable 180
Prevalence 3, 5–7, 35, 37–39, 86, 137–138
  rate 58
Prevention 70–71
Preventive clinical trial 78
Primary cancer 36–37, 85–86
Primary infection 5, 8
Proapoptotic activities 443
Probability 7–9
Progesterone 428
Progesterone receptor 55–56
Progress review group (PRG) 69
Prolactin response elements 428
Prospective cohort study 88–89, 218–219
Prostate cancer 13–14, 71, 110–111, 362–363
Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening trial 77
Prostatic fluid 392
Proteasome 443
Protein
  marker 198–203
  microarray 208
Proteome 198
Proteomics 197–198
  proteomic approaches 197–198

PSA (Prostate Specific Antigen) 11, 22–23
Public health research 70

Q

Quality of life 469–470

R

Race 3, 65–68, 227–235
  racial and ethnic disparities 68
  racial and ethnic variations 74, 228–230
  racial/ethnic minorities 70–76, 79
Radiation 96
Radiosensitive tissue 96
Radiosensitivity 376
Radiotherapy 96
Randomized clinical trials 9, 265
Randomized control trials 58, 179–180
Reactive oxygen species 291–292
Record keeping 230
Recurrence 461
Redox regulation 373
Registries 34
Regression 179–180
Relative risk 88–90, 240
Religious barriers 60
Reproductive factors 94
Restriction fragment length polymorphism (RFLP) 292–293
Retinoblastoma (Rb) 13–14, 96, 443–444
Retinoic acid 338–339
Retrospective cohort studies 218–219
Reverse phase microarray 208
Risk factors 3, 73–74, 90, 338–339, 397–398, 439–440
Risk pattern 92–93
Risk ratio 240

S

Saliva 392
Sample size 256–258
SAS 266
Screening 73, 86, 107–109, 221, 440, 488
  programs 57
Second primary malignancy 95–96
Secondary data 228
SEER (Surveillance, Epidemiology and End Results) 4, 32, 68, 74, 85–86
SEER Medicare-linked Data 34–35
Semen 392
Seroprevalence 392–393
Serostatus 393
Serum 199
Serum oestradiol levels (SEL) 265–266
Sexual orientation 66
Sigmoidoscopy 111
Single nucleotide polymorphism (SNP) 241
Small RNA profiling 273–274
Smoking 15–17
Smoking cessation 472
Social science data analysis network 67
Socio-economically disadvantaged 75
Socioeconomic status (SES) 65, 69, 74, 79, 233–234
Somatic mutations 474–475
Spindle cells 391, 397
Sporadic cancer 164–166
Squamous cell carcinoma (SCC) 86–87, 202, 335–336
Standardized incidence ratio (SIR) 89, 142, 219–220, 413
Standardized mortality ratio (SMR) 142, 240
Statistical analysis 32
Statistical efficiency 248
Statistical methods 239
Statistical techniques 36
Stomach cancer 17–18, 227–228
Stratification for variables 326
Study design 171–173
Suboptimal manual hormone receptor assay 55–56
Sumoylation 280–281
Surveillance 32, 138–139
Survival 3, 35, 65, 73–74, 85–86, 469
Synchronous 86–87

T

Tamoxifen 56, 97
Tangible factors 68
Tat proteins 388–389
Telemammography 111
Telephone interviews 199
Temporal trends 22–27, 35–36, 39, 41, 141–142
Testicular cancer 14
Testicular lymphoma 199–201
Therapy 234
Thyroid cancer 15
Tiling array 286–287
Tobacco 23–24, 338
  smoking 164, 337–338
Transmission 392
Transplant recipient 410
Transurethral resection 22–23
Treatment 58, 388
  outcome 32
Trial arm 479
Tryptic digest 204
Tumor 4, 33
  banking 99
  size 59
  suppressor 399, 457

Tumorigenesis 457–458
Two dimensional polyacrylamide gel electrophoresis (2 D-PAGE) 198

U

Ubiquitin protein 443
Ubiquitination 280
Unobserved variables (confounders) 169
Upper aerodigetive tract cancer 92
Urinary bladder cancer 14
Urinary tract cancer 86
Urothelial cell carcinoma 86

V

Vagina 95
Vascular endothelial growth factor (VEGF) 475
Vegetable consumption 93
Viremia 390
Viruses 421–434
  genome 389
  regulatory proteins 443
  viral factors 221
  virion size 440
Vitamin D 88
Vulva 95

W

Whites 66–67
World Health Organization (WHO) 69, 144, 439

X

X-ray repair cross-complementing 1 (XRCC1) 306
Xeroderma pigmentosum complementary group D (XPD) 306