

BIOMARKERS IN BREAST CANCER

CANCER DRUG DISCOVERY AND DEVELOPMENT

BEVERLY A. TEICHER, SERIES EDITOR

Cancer Drug Resistance, edited by *Beverly A. Teicher*, 2006

Histone Deacetylases: Transcriptional Regulation and Other Cellular Functions, edited by *Eric Verdin*, 2006

Immunotherapy of Cancer, edited by *Mary L. Disis*, 2006

Biomarkers in Breast Cancer: Molecular Diagnostics for Predicting and Monitoring Therapeutic Effect, edited by *Giampietro Gasparini and Daniel F. Hayes*, 2005

Protein Tyrosine Kinases: From Inhibitors to Useful Drugs, edited by *Doriana Fabbro and Frank McCormick*, 2005

Death Receptors in Cancer Therapy, edited by *Wafik S. El-Deiry*, 2005

Bone Metastasis: Experimental and Clinical Therapeutics, edited by *Gurmit Singh and Shafaat A. Rabbani*, 2005

The Oncogenomics Handbook, edited by *William J. LaRoche and Richard A. Shimkets*, 2005

Camptothecins in Cancer Therapy, edited by *Thomas G. Burke and Val R. Adams*, 2005

Combination Cancer Therapy: Modulators and Potentiators, edited by *Gary K. Schwartz*, 2005

Death Receptors in Cancer Therapy, edited by *Wafik S. El-Deiry*, 2005

Cancer Chemoprevention, Volume 2: Strategies for Cancer Chemoprevention, edited by *Gary J. Kelloff, Ernest T. Hawk, and Caroline C. Sigman*, 2005

Proteasome Inhibitors in Cancer Therapy, edited by *Julian Adams*, 2004

Nucleic Acid Therapeutics in Cancer, edited by *Alan M. Gewirtz*, 2004

Cancer Chemoprevention, Volume 1: Promising Cancer Chemopreventive Agents, edited by *Gary J. Kelloff, Ernest T. Hawk, and Caroline C. Sigman*, 2004

DNA Repair in Cancer Therapy, edited by *Lawrence C. Panasci and Moulay A. Alaoui-Jamali*, 2004

Hematopoietic Growth Factors in Oncology: Basic Science and Clinical Therapeutics, edited by *George Morstyn, MaryAnn Foote, and Graham J. Lieschke*, 2004

Handbook of Anticancer Pharmacokinetics and Pharmacodynamics, edited by *William D. Figg and Howard L. McLeod*, 2004

Anticancer Drug Development Guide: Preclinical Screening, Clinical Trials, and Approval, Second Edition, edited by *Beverly A. Teicher and Paul A. Andrews*, 2004

Handbook of Cancer Vaccines, edited by *Michael A. Morse, Timothy M. Clay, and Kim H. Lyerly*, 2004

Drug Delivery Systems in Cancer Therapy, edited by *Dennis M. Brown*, 2003

Oncogene-Directed Therapies, edited by *Janusz Rak*, 2003

Cell Cycle Inhibitors in Cancer Therapy: Current Strategies, edited by *Antonio Giordano and Kenneth J. Soprano*, 2003

Fluoropyrimidines in Cancer Therapy, edited by *Yousef M. Rustum*, 2003

Chemoradiation in Cancer Therapy, edited by *Hak Choy*, 2003

Targets for Cancer Chemotherapy: Transcription Factors and Other Nuclear Proteins, edited by *Nicholas B. La Thangue and Lan R. Bandara*, 2002

Tumor Targeting in Cancer Therapy, edited by *Michel Pagé*, 2002

Hormone Therapy in Breast and Prostate Cancer, edited by *V. Craig Jordan and Barrington J. A. Furr*, 2002

Tumor Models in Cancer Research, edited by *Beverly A. Teicher*, 2002

Tumor Suppressor Genes in Human Cancer, edited by *David E. Fisher*, 2001

Matrix Metalloproteinase Inhibitors in Cancer Therapy, edited by *Neil J. Clendeninn and Krzysztof Appelt*, 2001

Farnesyltransferase Inhibitors in Cancer, edited by *Said M. Sebti and Andrew D. Hamilton*, 2001

Platinum-Based Drugs in Cancer Therapy, edited by *Lloyd R. Kelland and Nicholas P. Farrell*, 2000

Signaling Networks and Cell Cycle Control: The Molecular Basis of Cancer and Other Diseases, edited by *J. Silvio Gutkind*, 1999

Apoptosis and Cancer Chemotherapy, edited by *John A. Hickman and Caroline Dive*, 1999

BIOMARKERS IN BREAST CANCER

*MOLECULAR DIAGNOSTICS FOR PREDICTING
AND MONITORING THERAPEUTIC EFFECT*

Edited by

GIAMPIETRO GASPARINI, MD

*Division of Medical Oncology
"San Filippo Neri" Hospital
Rome, Italy*

and

DANIEL F. HAYES, MD

*Breast Oncology Program
University of Michigan Comprehensive Cancer Center
Ann Arbor, MI*



HUMANA PRESS
TOTOWA, NEW JERSEY

© 2006 Humana Press Inc.
999 Riverview Drive, Suite 208
Totowa, New Jersey 07512
www.humanapress.com

All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording, or otherwise without written permission from the Publisher.

The content and opinions expressed in this book are the sole work of the authors and editors, who have warranted due diligence in the creation and issuance of their work. The publisher, editors, and authors are not responsible for errors or omissions or for any consequence arising from the information or opinions presented in this book and make no warranty, express or implied, with respect to its contents.

Due diligence has been taken by the publishers, editors, and authors of this book to assure the accuracy of the information published and to describe generally accepted practices. The contributors herein have carefully checked to ensure that the drug selections and dosages set forth in this text are accurate and in accord with the standards accepted at the time of publication. Notwithstanding, as new research, changes in government regulations, and knowledge from clinical experience relating to drug therapy and drug reactions constantly occurs, the reader is advised to check the product information provided by the manufacturer of each drug for any change in dosages or for additional warnings and contraindications. This is of utmost importance when the recommended drug herein is a new or infrequently used drug. It is the responsibility of the treating physician to determine dosages and treatment strategies for individual patients. Further it is the responsibility of the health care provider to ascertain the Food and Drug Administration status of each drug or device used in their clinical practice. The publisher, editors, and authors are not responsible for errors or omissions or for any consequences from the application of the information presented in this book and make no warranty, express or implied, with respect to the contents in this publication.

This publication is printed on acid-free paper. ∞

ANSI Z39.48-1984 (American National Standards Institute) Permanence of Paper for Printed Library Materials.

Cover design by Patricia F. Cleary.

For additional copies, pricing for bulk purchases, and/or information about other Humana titles, contact Humana at the above address or at any of the following numbers: Tel.: 973-256-1699; Fax: 973-256-8341; E-mail: orders@humanapr.com or visit our Website at www.humanapress.com

Photocopy Authorization Policy:

Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by Humana Press Inc., provided that the base fee of US \$30.00 per copy is paid directly to the Copyright Clearance Center at 222 Rosewood Drive, Danvers, MA 01923. For those organizations that have been granted a photocopy license from the CCC, a separate system of payment has been arranged and is acceptable to Humana Press Inc. The fee code for users of the Transactional Reporting Service is: [1-58829-227-4/06 \$30.00] .

Printed in the United States of America. 10 9 8 7 6 5 4 3 2 1

Library of Congress Cataloging-in-Publication Data

Biomarkers in breast cancer : molecular diagnostics for predicting and monitoring therapeutic effect / edited by Giampietro Gasparini and Daniel F. Hayes.

p. cm. — (Cancer drug discovery and development)

Includes bibliographical references and index.

ISBN 1-58829-227-4 (alk. paper) eISBN 1-59259-915-X

1. Breast—Cancer—Molecular diagnosis. 2. Breast—Cancer—Prognosis. 3. Tumor markers. I. Gasparini, Giampietro. II. Hayes, Daniel, 1951- III. Series.

RC280.B8B54 2005

616.99'449075—dc22

2005001922

PREFACE

Biomarkers in Breast Cancer: Molecular Diagnostics for Predicting and Monitoring Therapeutic Effect is an updated view of the prognostic and predictive biomarkers in breast cancer written by experts in this field. This book covers the major advancements in the application of novel sophisticated molecular methods as well as the state of the art of the conventional prognostic and predictive indicators.

The first three chapters by Simon, Sweep et al., and Kimel et al. highlight the relevance of appropriate and rigorous study design and guidelines for validation studies on new biomarkers, concerning the standardization with quality control of the assay(s) used for their determination, their clinical development, and the statistical approaches. Of particular importance is the suggested optimized protocol for the HER-2/*neu* FISH assay applied by the NSABP network (1).

Gene expression profiling by tissue microarray is treated in depth in the following two chapters by De Bortoli and Briglia and by Kim and Paik. Recent studies conducted using this methodology have clearly documented the heterogeneous nature of invasive breast cancer within the same pathologic stage and menopausal status. These methods have clearly provided powerful new tools for more accurate individual definition of prognosis. However, much more work remains to be done before standard pathology laboratories can use tissue microarrays to perform tumor marker studies for routine clinical use.

Several individual factors are accepted or appear to be promising for standard clinical care. Of course, the use of estrogen and progesterone receptors (ER, PgR) to predict benefit from endocrine therapy represents the gold standard of tumor markers, and should be tested on every breast cancer tissue (2). More recently, testing for HER-2 status has also become standard to help select whether a patient with metastatic breast cancer should receive trastuzumab (2). Other factors that are still controversial, but are considered standard by some guideline panels, include indicators of cell proliferation as well as the urokinase-type plasminogen activator (PAI)-1 system. These are the topics of the next four chapters written by authors who are among the pioneers and worldwide experts in translational research studies in the field.

The roles of the epidermal growth factor receptor (EGFR) pathway and altered p53 in breast cancer growth and progression and as possible prognostic and predictive indicators are outlined by Ciardiello et al. and by Kandoler and Jakesz, respectively. In addition to trastuzumab, several preclinical and clinical studies suggest that other agents that disrupt the signaling pathways generated by members of the EGFR family may be effective against breast cancer. The ability to identify predictive surrogate biomarkers of response is the key for the rational selection of the patients most likely to benefit from these anti-EGFR compounds.

Moreover, it seems likely that markers of EGFR activity might provide the opportunity for monitoring therapeutic efficacy. The recent demonstration that mutation of the phosphorylation site of the receptor may be predictive of gefitinib activity in patients with non-small-cell lung cancer may represent an important step toward the right strategy for the use of such a class of new anticancer agents (3,4).

Adjuvant systemic therapy has now been clearly shown to reduce the odds of recurrence and death (5,6). However, because of the inaccuracy of currently available prognostic and predictive factors, much of adjuvant systemic therapy, especially chemotherapy, is given inefficiently, either to patients whose cancer was never destined to recur or to those patients whose cancer will recur, but for whom the therapy will not be effective. Enormous work has been directed toward prognosis, with advancements in the field of molecular biology and in the detection of occult metastatic cells distant from the tumor. Hawes et al. and Braun et al. comprehensively cover the methodology, state of the art, pitfalls, and promises of detection of early tumor cell dissemination in breast cancer patients.

Although determination of tumor biology in tissue justifiably garners much interest, the ability to test and monitor for biological changes with a simple blood test is obviously appealing. Chapters 13 and 14 deal with the clinical significance of circulating HER-2/neu and vascular endothelial growth factor (VEGF), these being among the more promising therapeutic targets for approaches based on selective molecular-targeting agents.

We hope this text offers a critical view of the modern approach to the development of surrogate biomarkers of prognosis and responsiveness to selective treatments in breast cancer. We are confident that it will provide useful reading for investigators involved either in laboratory research or in clinical development of prognostic/predictive indicators and of novel molecular-targeted therapy.

Giampietro Gasparini, MD
Daniel F. Hayes, MD

1. Paik S, Bryant J, Tan-Chiu E, et al. Real-world performance of HER2 testing—National Surgical Adjuvant Breast and Bowel Project experience. *J Natl Cancer Inst* 2002;94:852–854.
2. Bast RC, Jr, Ravdin P, Hayes DF, et al. 2000 update of recommendations for the use of tumor markers in breast and colorectal cancer: clinical practice guidelines of the American Society of Clinical Oncology. *J Clin Oncol* 2001;19:1865–1878. Erratum in: *J Clin Oncol* 2001; 19:4185–4188. *J Clin Oncol* 2002;20:2213.
3. Lynch T, Bell D, Sordella R, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small cell lung cancer to gefitinib. *N Engl J Med* 2004;350:2129–2139.
4. Paez JG, Janne P, Lee JC, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 2004;304:1497–1500.
5. Early Breast Cancer Trialists' Collaborative Group. Tamoxifen for early breast cancer: an overview of the randomised trials. *Lancet* 1998;351:1451–1467.
6. Early Breast Cancer Trialists' Collaborative Group. Polychemotherapy for early breast cancer: an overview of the randomized trials. *Lancet* 1998;352:930–942.

CONTENTS

Preface	v
Contributors	ix

Part I. General Topics

- 1 Guidelines for the Design of Clinical Studies for the Development and Validation of Therapeutically Relevant Biomarkers and Biomarker-Based Classification Systems 3
Richard M. Simon
- 2 Analytical Aspects of Biomarker Immunoassays in Cancer Research 17
Fred C. G. J. Sweep, Chris M. G. Thomas, and Manfred Schmitt
- 3 Tissue Microarrays: *State of the Art, Pitfalls, and Promises* 31
Chungyeul Kim and Soonmyung Paik

Part II. Assays for Gene Expression and Post-Translational Protein Abnormalities

- 4 Gene Expression Profiling with DNA Microarrays: *Revolutionary Tools to Help Diagnosis, Prognosis, Treatment Guidance, and Drug Discovery* 47
Michele De Bortoli and Nicoletta Biglia

Part III. Assays for DNA Abnormalities

- 5 Assays for Gene Amplification: *State of the Art, Pitfalls, and Promises* 65
Chungyeul Kim, Yongkuk Song, and Soonmyung Paik

Part IV. Tissue Predictive Biomarkers

- 6 Cell Kinetics 81
Maria Grazia Daidone, Rosella Silvestrini, and Dino Amadori
- 7 Urokinase-Type Plasminogen Activator and PAI-1: *Validated Prognostic Factors for Breast Cancer* 111
Michael J. Duffy

8	Predictive Value of c- <i>erb</i> -B2 for Endocrine Therapy and Chemotherapy in Breast Cancer	129
	Vered Stearns	
9	<i>erb</i> -B2 as a Therapeutic Target	159
	Robert Mass	
10	The Epidermal Growth Factor Receptor Pathway as a Selective Molecular-Targeted Treatment in Human Breast Cancer	177
	Fortunato Ciardiello, Teresa Troiani, and Giampaolo Tortora	
11	<i>p53</i> as a Prognostic and Predictive Indicator	193
	Daniela Kandioler and Raimund Jakesz	
 Part V. Predictive Biomarkers in Body Fluids		
12	Occult Metastatic Cells in Breast Cancer Patients: <i>State of the Art, Pitfalls, and Promises</i>	213
	Stephan Braun, Julia Seeber, and Christian Marth	
13	Circulating HER2/neu: <i>Clinical Utility</i>	235
	Allan Lipton, Laurence Demers, Kim Leitzel, Suhail M. Ali, Rainer Neumann, Christopher P. Price, and Walter P. Carney	
14	Circulating Vascular Endothelial Growth Factor: <i>Methods, Prognostic Significance, and Potential Application for Antiangiogenic Therapy</i>	267
	Roberta Sarmiento, Roberta Franceschini, Sabrina Meo, Massimo Gion, Raffele Longo, and Giampietro Gasparini	
15	Detection of Early Tumor Dissemination in Patients with Breast Cancer	293
	Debra Hawes and Richard J. Cote	
 Part VI. Future Directions		
16	New Technologies/New Markers/New Challenges	325
	Daniel F. Hayes and Giampietro Gasparini	
Index		333

CONTRIBUTORS

- SUHAIL M. ALI, MD • *Medical Oncology Department, Penn State Milton S. Hershey Medical Center, Hershey, PA, USA*
- DINO AMADORI, MD • *Division of Oncology and Diagnostics, Pierantoni Hospital, Forlì, Italy*
- NICOLETTA BIGLIA, MD • *Unit of Gynecological Oncology, Institute for Cancer Research and Treatment (IRCC), University of Turin, Turin, Italy*
- STEPHAN BRAUN, MD • *Breast Center Tirol, Department of Obstetrics and Gynecology, University Hospital, Innsbruck, Austria*
- WALTER P. CARNEY, PhD • *Diagnostics Division, Oncogene Science/Bayer HealthCare, Cambridge, MA, USA*
- FORTUNATO CIARDIELLO, MD • *Cattedra di Oncologia Medica, Dipartimento Medico-Chirurgico di Internistica Clinica e Sperimentale F Magrassi e A Lanzara, Seconda Università degli Studi di Napoli, Naples, Italy*
- RICHARD J. COTE, MD, FRCPath • *Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA*
- MARIA GRAZIA DAIDONE, PhD • *Department of Experimental Oncology, National Cancer Institute, Milan, Italy*
- MICHELE DE BORTOLI, PhD • *Department of Oncological Sciences, Institute for Cancer Research and Treatment (IRCC), University of Turin, Turin, Italy*
- LAURENCE DEMERS, PhD • *Pathology Department, Penn State Milton S. Hershey Medical Center, Hershey, PA, USA*
- MICHAEL J. DUFFY, PhD, FRCPath, FACB • *Department of Nuclear Medicine, St. Vincent's Hospital, Dublin, Ireland*
- ROBERTA FRANCESCHINI, MS • *ABO Association c/o Centre for the Study of Biological Markers of Malignancy, Unit of Laboratory Medicine, Regional General Hospital, Venice, Italy*
- GIAMPIETRO GASPARINI, MD • *Division of Medical Oncology, San Filippo Neri Hospital, Rome, Italy*
- MASSIMO GION, MD • *Centre for the Study of Biological Markers of Malignancy, Unit of Laboratory Medicine, Regional General Hospital, Venice, Italy*
- DEBRA HAWES, MD • *Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA*

- DANIEL F. HAYES, MD • *Breast Oncology Program, University of Michigan Comprehensive Cancer Center, Department of Medicine, University of Michigan Health System, Ann Arbor, MI, USA*
- RAIMUND JAKESZ, MD • *The Department of Surgery, University of Vienna Medical School, Vienna, Austria*
- DANIELA KANDIOLER, MD • *The Department of Surgery, University of Vienna Medical School, Vienna, Austria*
- CHUNGYEUL KIM, MD • *Division of Pathology, National Surgical Adjuvant Breast and Bowel Project Foundation, Pittsburgh, PA, USA*
- KIM LEITZEL, MS • *Medical Oncology Department, Penn State Milton S. Hershey Medical Center, Hershey, PA, USA*
- ALLAN LIPTON, MD • *Medical Oncology Department, Penn State Milton S. Hershey Medical Center, Hershey, PA, USA*
- RAFFELE LONGO, MD • *Division of Medical Oncology, San Filippo Neri Hospital, Rome, Italy*
- CHRISTIAN MARTH, MD • *Breast Center Tirol, Department of Obstetrics and Gynecology, University Hospital, Innsbruck, Austria*
- ROBERT MASS, MD • *Clinical Scientist, Genentech, Inc., South San Francisco, CA, USA*
- SABRINA MEO, MS • *ABO Association c/o Centre for the Study of Biological Markers of Malignancy, Unit of Laboratory Medicine, Regional General Hospital, Venice, Italy*
- RAINER NEUMANN, PhD • *Bayer HealthCare, Leverkusen, Germany*
- SOONMYUNG PAIK, MD • *Division of Pathology, National Surgical Adjuvant Breast and Bowel Project Foundation, Pittsburgh, PA, USA*
- CHRISTOPHER P. PRICE, PhD • *Diagnostics Division, Bayer Health Care, Slough, United Kingdom*
- ROBERTA SARMIENTO, MD • *Division of Medical Oncology, San Filippo Neri Hospital, Rome, Italy*
- MANFRED SCHMITT, PhD • *Clinical Research Unit, Department of Gynecology, Technical University of Munich, Munich, Germany*
- JULIA SEEBER, MD • *Breast Center Tirol, Department of Obstetrics and Gynecology, University Hospital, Innsbruck, Austria*
- ROSELLA SILVESTRINI, PhD • *National Cancer Institute, Milan, Italy*
- RICHARD M. SIMON, DSc • *Biometric Research Branch, National Cancer Institute, Bethesda, MD, USA*
- YONGKUK SONG, PhD • *Division of Pathology, National Surgical Adjuvant Breast and Bowel Project Foundation, Pittsburgh, PA, USA*

VERED STEARNS, MD • *The Sidney Kimmel Comprehensive Cancer Center,
Johns Hopkins University, Baltimore, MD, USA*

FRED C. G. J. SWEEP, PhD • *Department of Chemical Endocrinology, University
Medical Centre Nijmegen, Nijmegen, The Netherlands*

CHRIS M. G. THOMAS, PhD • *Department of Chemical Endocrinology, University
Medical Centre Nijmegen, Nijmegen, The Netherlands*

GIAMPAOLO TORTORA, MD • *Division of Medical Oncology, Department of Molecular
and Clinical Endocrinology and Oncology, Università di Napoli Federico II,
Naples, Italy*

TERESA TROIANI, MD • *Division of Medical Oncology, Department of Molecular
and Clinical Endocrinology and Oncology, Università di Napoli Federico II,
Naples, Italy*

I

GENERAL TOPICS

1

Guidelines for the Design of Clinical Studies for the Development and Validation of Therapeutically Relevant Biomarkers and Biomarker- Based Classification Systems

Richard M. Simon, DSc

CONTENTS

INTRODUCTION
PITFALLS IN DEVELOPMENTAL STUDIES
STRUCTURED RETROSPECTIVE STUDIES
VALIDATION STUDIES
REFERENCES

SUMMARY

Standards for the development of therapeutically relevant biomarkers and biomarker-based classification systems are lacking. The literature of prognostic marker studies for breast cancer is inconsistent, and few such markers have been adopted for widespread use in clinical practice. This is problematic, as many patients are over-treated and many others are treated ineffectively. The deficiencies in clinical development of biomarkers may become more severe as DNA

*Cancer Drug Discovery and Development: Biomarkers in Breast Cancer:
Molecular Diagnostics for Predicting and Monitoring Therapeutic Effect*

Edited by: G. Gasparini and D. F. Hayes © Humana Press Inc., Totowa, NJ

microarrays and proteomic technologies provide many new candidate markers and therapeutics become more molecularly targeted. In this chapter we address some common problems with developmental marker studies and provide recommendations for the design of clinical studies for the development and validation of robust, reproducible, and therapeutically relevant biomarkers and biomarker-based classification systems. The design of validation studies is addressed for (1) identifying node-negative breast cancer patients who do not require systemic chemotherapy; (2) identifying node-positive breast cancer patients who do not benefit from standard chemotherapy; and (3) identifying node-positive breast cancer patients who benefit from a new molecularly targeted therapeutic.

Key Words: Biomarkers; microarrays; classification systems; clinical trial design.

1. INTRODUCTION

Breast cancer is a heterogeneous set of diseases. Although substantial progress has been made in the treatment of breast cancer, many patients are overtreated and many undergo intensive chemotherapy with little apparent benefit. The literature on prognostic factors in breast cancer, although voluminous, is inconsistent (1). The process of how to develop biomarkers that are robust, reproducibly measured, and therapeutically effective has not been well established. Although many prognostic factors have been studied, treatment selection has remained based primarily on the traditional components of Tumor–Node–Metastasis (TNM) stage and hormone receptor levels. This discrepancy between an inconsistent research literature and clinical practice will become even more problematic as DNA microarray and proteomic technologies provide new markers and therapeutics become more molecularly targeted. The objectives of this chapter are to provide information that facilitates the development of biomarkers for selection of the best treatment for each patient. We use the term *biomarker* to include predictive classification systems based on protein or RNA transcript profiles measured using technology such as DNA microarrays.

2. PITFALLS IN DEVELOPMENTAL STUDIES

Most biomarkers are developed using archived tumor specimens, and many of the problems that exist in the marker literature derive from the retrospective nature of these studies. Clinical drug trials are generally prospective, with patient selection criteria, primary end point, hypotheses, and analysis plan specified in advance in a written protocol. The consumers of clinical trial reports have been educated to be skeptical of *data dredging* to

find something “statistically significant” to report in clinical trials. They are skeptical of analyses with multiple end points or multiple subsets, knowing that the chances of erroneous conclusions increase rapidly once one leaves the context of a focused single-hypothesis clinical trial. Marker studies are generally performed with no written protocol, no eligibility criteria, no primary end point or hypotheses, and no defined analysis plan. The patient population is often very heterogeneous and represents individuals for whom archived specimens are available. The patients are often not treated in a single clinical trial and represent a mixture of stages. Consequently, the overall population often does not represent a therapeutically meaningful group and the biomarkers identified may be of prognostic relevance, but less likely to be of predictive relevance for selecting therapy. Often the marker may be prognostic because it is correlated with disease stage or some other known prognostic marker. Broad populations are also often heterogeneously treated and so finding that a marker is prognostic in such a population may be difficult to interpret. Prognostic markers that do not have therapeutic implications are rarely used. The heterogeneous nature of the population also often results in multiple subset analyses of more therapeutically meaningful subpopulations. With multiple analyses, the chance of false-positive conclusions increases. Many biomarker studies perform analyses for many candidate biomarkers and several end points as well as for various patient subsets. Consequently, the chance for erroneous conclusions increases multiplicatively. The multiplicity problem is even more severe when one considers that there are usually multiple ways of quantifying biomarker level and many possible mathematical models for combining biomarker measurements.

Many of the problems that have hindered the development and acceptance of predictive single-protein biomarkers also apply for biomarkers based on DNA microarray expression profiles (2). There are multiple platforms and protocols for measuring expression profiles, and microarray research studies almost never evaluate interlaboratory assay reproducibility. Microarray expression profiles in research studies are generally performed at one time so that reagent variability is minimized, and it is almost never demonstrated that the models are predictive for tumor specimens collected and assayed at other times. This is of particular concern for printed cDNA microarrays where there may be substantial variability among batches of printed slides and batches of reference RNA.

Because of the number of genes available for analysis, microarray data can be a veritable fountain of false findings unless appropriate statistical methods are utilized. For example, in comparing expression profiles of 10,000 genes for tumor specimens selected from patients who have responded to a specified treatment to those for nonresponders, the expected number of false-positive genes that are statistically significantly ($p < 0.05$)

differentially expressed between the two groups is 500. This is true regardless of whether the expression levels for different genes are correlated. Consequently, more stringent methods for assessing differential expression must be used. Some studies do not use statistical significance at all and just identify genes as differentially expressed based on fold-change statistics; that is, the ratio of the average expression level in responders to the average in nonresponders, ignoring variability entirely. Others base their analyses on visual inspection of graphical data displays. Such methods are clearly problematic.

The unstructured nature of retrospective studies of biomarkers would not be so problematic if they were followed by structured prospective validation studies that tested specific hypotheses about predictive biomarkers. Such prospective trials are rarely performed, however, because they are difficult to accomplish. Consequently, before discussing the design of such prospective trials, we will make some suggestions about a more structured approach to retrospective studies.

3. STRUCTURED RETROSPECTIVE STUDIES

There is a role for exploratory studies in which multiple biomarkers and multiple ways of combining biomarkers into predictive models are examined so long as one has an adequate way of evaluating the result. A major problem with many retrospective studies is that they attempt to use the same set of data to both develop hypotheses (biomarkers) and to test those hypotheses. This problem is particularly severe when the number of candidate hypotheses examined in the exploratory stage is large.

In trying to determine which genes are differentially expressed in comparing responders to a given therapy to nonresponders, the number of hypotheses equals the number of genes examined. The Bonferonni method of adjusting for multiple testing requires that the p value calculated for comparing expression of a specific gene i in responders to nonresponders, say p_i , be adjusted based on the number of genes (N) examined. For microarray studies, N could be 10,000 or greater. The Bonferonni method tries to eliminate all false positives. For microarray studies, less conservative methods control the number of *false discoveries* (false positives), or the proportion of claimed positives that are false positives (*false discovery rate*) (3). These same ideas apply if, for example, we are examining which genes are prognostic for survival or disease-free survival on a particular treatment.

For assessing statistical significance, adjustments such as those described in the preceding paragraph can be applied to adjust for the fact that we do not have a specific hypothesis to test, but rather are in a hypothesis development mode. The adjustment is based on treating the problem as one of testing all possible hypotheses. For retrospective biomarker studies in which

a number of biomarkers are examined, such adjustments to statistical significance should be applied. In many cases, however, statistical significance is not the best measure of biomarker value. A better measure is the extent to which the biomarker model enables us to predict whether the patient will respond to the treatment (4).

For binary outcomes such as response and nonresponse, the best measure of predictive accuracy is the number of correct predictions. For quantitative outcomes such as survival or disease-free survival, measurement of predictive accuracy is more complex. In many cases, it is reasonable to approximate quantitative outcomes in a binary manner: good outcome or poor outcome. In other cases, measures such as described by Korn and Simon (5) are used.

It is not valid to use the same set of data for selecting a predictive marker or developing a predictive model and for measuring predictive accuracy. The estimate of predictive accuracy computed on the same data used to select the marker or develop the model is called the *resubstitution* estimate and is known to be biased (6). The bias is extreme when the number of candidate markers is larger than the number of cases. For example, Simon et al. (6) showed that for two classes (e.g., responders and nonresponders) that have no genes that are truly differentially expressed in microarray expression profiles of thousands of genes, one can almost always find a predictive model that has a resubstitution estimate of accuracy of 100%. Such a model would be useless for future data, but would appear to give perfect predictions for the cases used to develop the model.

How can we develop a proper estimate of the accuracy of class prediction for future samples? For a future sample, we will apply a fully specified predictor developed using the data available today. If we are to emulate the future predictive setting in developing our estimate of predictive accuracy, we must set aside some of our samples and make them completely inaccessible until we have a fully specified predictor that has been developed from scratch without utilizing those set-aside samples.

To estimate properly the accuracy of a predictor for future samples, the current set of samples must be partitioned into a training set and a separate test set. The test set emulates the set of future samples for which class labels are to be predicted. Consequently the test samples cannot be used in any way for the development of the prediction model. This means that the test samples cannot be used for estimating the parameters of the model and they cannot be used for selecting the gene set to be used in the model. It is this latter point that is often overlooked.

The most straightforward method of estimating the accuracy of future prediction is the *split-sample* method of partitioning the set of samples into a training set and a test set as described in the previous paragraph. Rosenwald et al. (7) used this approach successfully in their international study of

prognostic prediction for large cell lymphoma. They used two thirds of their samples as a training set. Multiple kinds of predictors were studied on the training set. When the collaborators of that study agreed on a single fully specified prediction model, they accessed the test set for the first time. On the test set there was no adjustment of the model, redefining of cutoff values, or fitting of parameters. They merely used the samples in the test set to evaluate the predictions of the model that was completely specified using only the training data.

Cross-validation is an alternative to the split sample method of estimating prediction accuracy (8). Cross-validation can be used only when there is a well-defined algorithm for predictive model development. In such cases, cross-validation can be more efficient than the split-sample method for estimating prediction accuracy. There are several forms of cross-validation. Here we will describe *leave-one-out cross-validation (LOOCV)* in the context of a class predictor based on gene expression levels determined by DNA microarray analysis. LOOCV starts like split-sample cross-validation in forming a training set of samples and a test set. With LOOCV, however, the test set consists of only a single sample; the rest of the samples are placed in the training set. The sample in the test set is placed aside and not utilized at all in the development of the class prediction model. Using only the training set, the informative genes are selected and the parameters of the model are fit to the data. Let us call M_1 the model developed with sample 1 in the test set. When this model is fully developed, it is used to predict the class of sample 1. This prediction is made using the expression profile of sample 1, but obviously without using knowledge of the true class of sample 1. Symbolically, if \underline{x}_1 denotes the complete expression profile of sample 1, then we apply model M_1 to \underline{x}_1 to obtain a predicted class \hat{c}_1 . This predicted class is compared to the true class label c_1 of sample 1. If they disagree, then the prediction is in error. Then a new training set–test set partition is created. This time sample 2 is placed in the test set and all of the other samples, including sample 1, are placed in the training set. A new model is constructed from scratch using the samples in the new training set. Call this model M_2 . Model M_2 will generally not contain the same genes as model M_1 . Although the same algorithm for gene selection and parameter estimation is used, since model M_2 is constructed from scratch on the new training set, it will in general not contain exactly the same gene set as M_1 . After creating M_2 , it is applied to the expression profile \underline{x}_2 of the sample in the new test set to obtain a predicted class \hat{c}_2 . If this predicted class does not agree with the true class label c_2 of the second sample, then the prediction is in error.

The process described in the previous paragraph is repeated n times, where n is the number of biologically independent samples. Each time it is applied, a different sample is used to form the single-sample test set. During the steps, n different models are created and each one is used to predict the

class of the omitted sample. The number of prediction errors is totaled and reported as the leave-one-out cross-validated estimate of the prediction error.

At the end of the LOOCV procedure you have constructed n different models. They were constructed only in order to estimate the prediction error associated with the model constructed by applying the algorithm to the complete set of samples. The model that would be used for future predictions is one constructed using all n samples. That is the best model for future prediction and the one that should be reported in the publication. The cross-validated error rate is an estimate of the error rate to be expected in use of this model for future samples assuming that the relationship between class and expression profile is the same for future samples as for the currently available samples. With two classes, one can use a similar approach to obtain cross-validated estimates of the sensitivity, specificity.

Leave-one-out cross-validation is applicable only in settings in which there is an algorithm for the development of a predictive model. In many studies, the analysis is less algorithmic and many kinds of prediction models are explored. For such studies, it is best to use the split sample approach of setting aside at least one third of the samples as a test or validation set. The samples in the test set should not be used for any purpose other than testing the final model developed in the training set. Specifically, the test set samples should not be used for limiting the set of genes to be considered in detail in the training set. The samples in the test set should not be accessed until a single model is identified based on training set analyses as *the* model to be tested.

LOOCV can be used to evaluate risk group predictors using survival or disease-free survival data. Suppose we wish to identify patients in a low-risk group with 10-year disease-free survival $>90\%$. Consider the leave-one-out training set in which observation i is left out in the test set. A disease-free survival model M_i is developed for the training set. For example, the model might be a proportional hazards regression model that predicts disease-free survival based on the expression profile and/or standard prognostic factors. The model M_i is applied to the left-out specimen i to obtain a prediction of the probability that the i th patient has 10-yr disease-free survival $>90\%$. Let $y_i = 1$ if this probability is greater than 50%. This process is repeated for all of the leave-one-out training sets. Then, the Kaplan–Meier disease-free survival curve estimate is computed and plotted for the patients predicted to be of very low risk, those with $y_i = 1$. The adequacy of the model is judged by whether the estimated 10-yr disease-free survival for the identified low-risk group is in fact in excess of 90%. An approach similar to this was used for developing a classification system based on survival for patients with renal cancer by Vasselli et al. (9).

One of the common errors in retrospective studies of biomarkers is that the statistical significance of the biomarker is evaluated rather than the

predictive accuracy of the biomarker (4). We have indicated in the preceding how predictive accuracy can be evaluated in a manner that avoids the bias of the resubstitution estimate. But even this is not sufficient. New biomarkers are often correlated with existing prognostic factors. The retrospective study must provide strong evidence that the new marker is substantially more predictive than the currently available prognostic factors. This can be addressed by computing the split-sample or cross-validated error rate for a model consisting of current prognostic factors and then computing the split-sample or cross-validated error rate for a model consisting of current prognostic factors plus the new candidate markers. Only if the latter is substantially greater than the former with regard to a therapeutically relevant prediction will a prospective validation study be warranted.

4. VALIDATION STUDIES

Assuming that the initial study is performed properly with attention to the statistical principles described in previous section, it might be considered a phase II study, and the next step should be to conduct a phase III study that is focused on testing the specific classifier developed by the initial study (10). The phase III study should be conducted with a written protocol. The phase III trial should be designed to test the biomarker classifier developed in the previous study. The classifier should be fully specified in the protocol. If the biomarker is expression profile based, the specification must include the genes used, the mathematical form of the classifier, parameter values, and cutoff thresholds for distinguishing the classes or prognostic groups.

The phase III study should attempt to perform the assays in a manner as similar as possible to the way it would be performed broadly outside of a research setting if the diagnostic classifier were adopted. Consequently, attention is required in determining whether the same platform should be used for the phase III trial as for the phase II trial. If the platform is changed, then clearly some intermediate study will be needed to translate the classification algorithm from use on the phase II platform to the platform used in the phase III trial.

Even if there is not a change in platform, an intermediate study may be required to prepare the classifier for use with multiple laboratories performing the assay. In the phase II trial all of the assays may have been performed at a single location by a research laboratory and it may be advisable to conduct the phase III trial in a manner more similar to the way it would be performed if the classifier were adopted for national use. Generally this will mean that several laboratories will be conducting the assays. Consequently, the protocol for the phase III study should specify procedures to be used for conducting the assay. It is also useful to conduct intermediate studies of interlaboratory reproducibility of the assays. Unless interlaboratory repro-

Table 1
Guidelines for Validation Studies (10)

1. Intra- and interlaboratory reproducibility of assays should be documented.
2. Laboratory assays should be performed blinded to clinical data and outcome.
3. An inception cohort of patients should be assembled with <15% of patients nonevaluable owing to missing tissue or data. The referral pattern and eligibility criteria should be described.
4. Treatment should be standardized or randomized and accounted for in the analysis.
5. Hypotheses should be stated in advance, including specification of prognostic factors, coding of prognostic factors, end points, and subsets of patients and treatments.
6. The sample size and number of events should be sufficiently large that statistically reliable results are obtained. Statistical power calculations that incorporate the number of hypotheses to be tested and appropriate subsets for each hypothesis should be described. There should be at least 10 events per prognostic factor examined per subset analyzed.
7. Analyses should test whether new factors add predictiveness after adjustment for or within subsets determined by standard prognostic factors.
8. Analyses should be adjusted for the number of hypotheses to be tested.
9. Analyses should be based on prespecified cutoff values for prognostic factors or cutoffs should be avoided.

ducibility is sufficiently high, it is not advisable to proceed with the phase III trial.

If the biomarker classifier was developed using a dual-label microarray platform, then use of the classifier in other laboratories requires that they use the same common reference RNA as was used for the initial study. Because different batches of the common reference will be utilized for classifying subsequent patients, calibration studies will generally be required to ensure that the expression profile of the common reference does not change and to adjust the classifier for small changes.

Conducting the validation study as a prospective trial is desirable for many reasons. One can never be sure that the patients for whom one has adequate preserved tissue are representative of the population of patients presenting for treatment. It is difficult to assure that a retrospective cohort was adequately staged and treated, and the data available may be incomplete. It is also difficult to assess whether a diagnostic procedure is practical unless it is studied in the real-time context of presenting patients who need to be evaluated and treated. Prospective accrual is also important for evaluating the diagnostic classifier in the context of real-time tissue handling. Table 1, reprinted from Simon and Altman (10), indicates some important design features of prospective validation studies.

The objective of the validation trial of a predictive marker is to test the hypothesis that the marker is useful for treatment selection. This is often a more complex objective than validation of a prognostic marker, in which the objective is to determine whether the marker can separate the uniformly staged and treated patients into groups of differing outcome. There are some cases, discussed below, in which prognostic markers are also predictive markers. Our focus is on predictive markers and we will consider three breast cancer scenarios.

4.1. Identifying Node-Negative Patients Who Do Not Require Chemotherapy

Our first scenario is a putative marker for identifying node-negative patients whose prognosis on local therapy and possibly tamoxifen is so good that they do not require chemotherapy. The retrospective study for development of such a marker would have probably been based on archived tumors of node-negative patients who did not receive chemotherapy. A tissue microarray of a large number of such specimens, with associated clinical follow-up data, can provide a valuable resource for ensuring that the marker is sufficiently promising to warrant evaluation in a prospective clinical trial if the classifier is not RNA transcript profile based. A marker of this type meets the definition of a prognostic marker, but it can also be a predictive marker if it enables us to determine which node negative patients do not require chemotherapy.

The theoretically optimal trial design would be to randomize candidate node-negative patients to receive or not receive chemotherapy and then to validate whether the marker identifies those who do not benefit from chemotherapy. The candidate node-negative patients might be those with tumors 1–3 cm in diameter without known poor prognostic features such as hormone receptor negativity. This is probably not a feasible approach, however, because chemotherapy has already been established as being effective for much of the candidate population.

An alternative study design is to withhold chemotherapy from a subset of node-negative patients selected based on marker status to be of particularly low risk. If their outcomes were sufficiently good relative to some standard, then the marker would be accepted as useful. The standard might be based on outcomes for node-negative patients that are similar with regard to standard prognostic factors in other studies. It may also be useful to compare outcome for the selected patients (M^+) to the outcome for the patients of the same series who did not have such predicted good prognosis (M^-). The latter patients would have received chemotherapy, but their outcome even with chemotherapy may not be as good as that of the M^+ patients without chemotherapy. If that is the case, then the value of the marker for withholding chemotherapy will have been demonstrated.

An alternative approach would be to randomize patients selected as low risk based on marker status (M^+) to either receive or not receive chemotherapy. The marker would be validated if the randomized trial demonstrated that there was no clinically significant benefit of chemotherapy in the selected subset of patients. This would have to be a very large clinical trial, however. The benefit of chemotherapy would be expected to reduce the hazard or recurrence only by approx 25%, and with a very low event rate this is equivalent to a very small difference in absolute disease-free survival. The randomized trial asks a different question than the strategy described in the previous paragraph. The randomized trial asks whether there is a benefit of treatment. For patients with a very good prognosis, however, a statistically significant treatment effect may be of questionable clinical significance. Consequently, the randomized trial may not answer the most relevant question.

Gasparini et al. (11) describe guidelines for the adoption into clinical practice of new prognostic markers for use in treatment selection for patients with node-negative breast cancer.

4.2. Identifying Node-Positive Patients Who Do Not Benefit From a Chemotherapy Regimen

Consider now a putative marker that permits the identification of patients who do not benefit from a chemotherapy regimen that has been standard treatment. Let T denote the chemotherapy regimen, and let S denote local therapy or local therapy plus tamoxifen. The retrospective study used to develop the marker may have been based on tissue from a randomized trial of T vs S. In some cases the marker may be based on finding a signature of patients who do not respond to T in metastatic disease trials.

The ideal validation trial would probably be a randomized trial of S vs T for patients with node-positive breast cancer. One could analyze such a trial by seeing whether the benefit of T vs S depended on the marker level. Such a trial would generally be impractical, however, because T or some other kind of chemotherapy is standard treatment for node-positive patients. It might be possible, however, to randomize patients to receive or not receive one or more courses of T preoperatively, and to correlate marker result with biological response to T as assessed from the surgical specimen.

A second strategy would be to use chemotherapy T on all patients after measuring the marker. One could then determine prospectively whether the marker level correlates with outcome. This is a strategy analogous to that recommended in Subheading 4.1. Here, however, one is trying to determine whether the marker identifies a group of such poor disease-free survival on standard treatment T that the chemotherapy is judged nonworthwhile even in the absence of a control group not receiving chemotherapy. This strategy

may be less satisfactory for judging poor prognosis in absolute terms than it was in Subheading 4.1. for judging good prognosis.

A third strategy would be to randomize the patients to marker based vs non-marker-based therapeutic management. The non-marker-based management would assign T to all patients. The marker-based management would assign T to all except those predicted based on the marker to be nonresponsive (M^-). One way of conducting such a trial is to measure the marker only for those patients assigned to marker-based management. The value of the marker is determined by comparing the outcome for the marker-based management arm to the outcome for the non-marker-based management arm. This is, however, a very inefficient trial design. Because most patients in both arms of the trial will be receiving the same treatment, the average treatment difference will be very small between the arms and a huge sample size will be required. The situation is even more problematic because it is a therapeutic equivalence trial in the sense that failure to find a statistically significant difference leads to the adoption of the new treatment approach, in this case marker-based treatment assignment.

A better design is to measure the marker on all patients, and then randomize them to marker-based treatment vs non-marker-based treatment. The evaluation of the marker can be performed by comparing outcomes for the M^- patients who received chemotherapy T on the non-marker-based arm but treatment S on the marker-based arm. This will require a much smaller sample size than the design described in the previous paragraph. This design is essentially equivalent to randomizing the M^- patients to T or S.

4.3. Identifying Node-Positive Patients Who Benefit From a Specific Regimen

Our third scenario is that we have a putative marker that identifies patients whose tumors are responsive to a new regimen E when the standard chemotherapeutic regimen is T. Many new therapeutics have defined molecular targets and are developed in conjunction with an assay that measures the expression of the target. The most adequate validation study is often a randomized clinical trial in which both marker-positive and marker-negative patients are randomized to either standard treatment T or T plus the new regimen E. The trial should be large enough so that the new regimen can be evaluated separately in the M^+ and M^- subsets. This requires about twice as many patients as if the regimen T+E were to be evaluated overall, without reference to the marker.

If the biological relationship between the marker and the therapeutic is sufficiently strong, it may be difficult to justify including marker negative patients in the study. A randomized study comparing T to T+E for M^+ patients may be very efficient for demonstrating the effectiveness of the new treatment E, but it will not really constitute a validation of the marker.

The development of the therapeutic, supported by the marker assay, may, however, be more important than validation of the essentiality of the marker for selecting patients.

The least desirable alternative would be to randomize patients between T and T+E without measuring the marker. If the marker is important, then such a trial design may be very inefficient for evaluating the therapeutic E, and of course, it provides no information for validating the marker.

The scenario described here is also applicable to the development of treatment regimens in which the molecular target is not known or not known with certainty. Instead of using an assay based on the expression of the putative target, one may use a DNA microarray-based classifier developed in phase II trials of metastatic disease patients for distinguishing responders from nonresponders to the new regimen E. If tumor specimens are available from patients treated with the standard treatment T as well as those treated with the new treatment E, the classifier can be developed to identify patients who are predicted to be more responsive to the new treatment E but not to standard treatment T.

REFERENCES

1. Hilsenbeck SG, Clark GM, McGuire WL. Why do so many prognostic factors fail to pan out? *Breast Cancer Res Treat* 1992;22:197–206.
2. Simon RM, Korn EL, McShane LM, Radmacher MD, Wright GW, Zhao Y. *Design and analysis of DNA microarray investigations*. Springer, New York, 2003.
3. Reiner A, Yekutieli D, Benjamini Y. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* 2003;19:368–375.
4. Kattan MW. Judging new markers by their ability to improve predictive accuracy. *J Natl Cancer Inst* 2003;95:634–635.
5. Korn EL, Simon R. Measures of explained variation for survival data. *Stat Med* 1990;9:487–504.
6. Simon R, Radmacher MD, Dobbin K, McShane LM. Pitfalls in the analysis of DNA microarray data: class prediction methods. *J Natl Cancer Inst* 2003;95:14–18.
7. Rosenwald A, Wright G, Chan WC, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med* 2002;346:1937–1947.
8. Radmacher MD, McShane LM, Simon R. A paradigm for class prediction using gene expression profiles. *J Comput Biol* 2002;9:505–511.
9. Vasselli J, Shih JH, Iyengar SR, et al. Predicting survival in patients with metastatic kidney cancer by gene expression profiling in the primary tumor. *Proc Natl Acad Sci USA* 2003;100:6958–6963.
10. Simon R, Altman DG. Statistical aspects of prognostic factor studies in oncology. *Br J Cancer* 1994;69:979–985.
11. Gasparini G, Pozza F, Harris AL. Evaluating the potential usefulness of new prognostic and predictive indicators in node-negative breast cancer patients. *J Natl Cancer Inst* 1993;85:1206–1219.

2

Analytical Aspects of Biomarker Immunoassays in Cancer Research

*Fred C.G.J. Sweep, PhD,
Chris M.G. Thomas, PhD,
and Manfred Schmitt, PhD*

CONTENTS

INTRODUCTION
IMMUNOASSAY DESIGNS
VARIABILITY IN TEST RESULTS
ASPECTS OF BIOMARKER ASSESSMENT
QUALITY ASSURANCE
CONCLUSIONS
REFERENCES

SUMMARY

Many difficulties associated with immuno(metric) assay kits designed for quantification of a particular biomarker arise from their variation in specificity and binding affinity of the employed antibodies. Other important sources causing varying assay results are the use of different standard preparations in these kits and the nonuniform preanalytical specimen processing procedures employed, each of which should be subjected to standardization. To improve the performance and comparability of assays, continuous interlaboratory external quality control procedures are needed. Such quality assurance

*Cancer Drug Discovery and Development: Biomarkers in Breast Cancer:
Molecular Diagnostics for Predicting and Monitoring Therapeutic Effect*

Edited by: G. Gasparini and D. F. Hayes © Humana Press Inc., Totowa, NJ

programs provide a forum for expert laboratory investigators to discuss technical details and to exchange laboratory issues and related practical information. This chapter addresses some of these issues and presents initial analytical validation procedures of newly developed biomarker assays, the validation of already established assay procedures for routine use on a day-to-day basis, and finally discusses some aspects on adequate (external) quality control proficiency testing.

Key Words: Biomarkers; immunoassay; cancer; tumor markers.

1. INTRODUCTION

In many cases progression of cancer growth is rather slow and often it may take years for a malignancy to manifest clinically. Because early cancer detection is required to significantly reduce cancer mortality, screening procedures are needed that are highly specific (i.e., providing almost a 100% proportion of negative test results for a tumour marker in nondiseased individuals) and sensitive enough to detect malignancies at an early stage of development. Thus, the screening procedure should give assay results above a defined cutoff value in a reasonable proportion of early stage diseased persons. As yet, there are no assay procedures available that meet such a specification, although there is a growing public interest in improving early cancer detection. Ideally, determination in biological specimens of cancer-derived analytes for a particular type of cancer not only should provide valuable information for initial diagnosis, but also should have prognostic value to guide the choice of treatment, and such a test should provide a reflection of the tumor burden of the patient, being predictive for recurrent disease after initial treatment, and of help in monitoring the course of the disease throughout time of follow-up. Each of these properties should contribute to more effective treatment of an individual patient and thus provide indispensable information for improving the quality of life and outcome of the disease by increasing disease-free and overall survival. Despite extensive research efforts in the last decades and numerous papers dealing with development and clinical testing of potentially promising biochemical markers, no assays are as yet available that are sensitive enough to convincingly detect any of the major types of cancer at the most early stage. Although an impressive number of biochemical markers with the capacity to predict disease recurrence and/or early death have been introduced, a comprehensive understanding of the tumor biological processes involved is still lacking.

1.1. Guidelines for Evaluating Clinical Value of Biomarkers

At present, cancer diagnosis is based mainly on clinical symptoms and confirmed by histomorphological findings. Application of biochemical markers in this process may have additional value but still, depending on the

marker test applied, the reliability criteria may become less important. In case of screening and diagnosis (“rule-in-disease”) specificity is of utmost importance (to avoid false-positive assay results leading to unfavorable and unnecessary medical examination and treatment), although it should be realized that increasing specificity of a test goes at the cost of decreasing its sensitivity (that should remain high enough to detect early-stage diseased individuals). If the purpose of the test is disease monitoring to detect recurrence during follow-up, precision should be high; providing a prognosis for treatment, the test should put emphasis on specificity and accuracy. All of these criteria are not well established and should become standard criteria for evaluation of biomarker assays and their clinical application. In line with this, Hayes et al. (1) proposed certain criteria to standardize the available biomarker information for clinical use in a biomarker utility grading system.

Currently, many biochemical markers of potentially prognostic value are intensively tested in multicenter clinical trials. Only a few of these show a benefit for predicting prognosis of node-negative breast cancer patients. To conclude that these newly developed biochemical markers have independent prognostic value over already known factors, McGuire and Clark (2) some years ago proposed strict guidelines for evaluating newly developed prognostic markers, addressing the biological role of the new factor as well as the extent of the sample size, the risk of sample bias, the appropriate testing system, the establishment of cutoff values in a training data set, and confirmation of these observations in an independent validation data set.

2. IMMUNOASSAY DESIGNS

Immunochemical assay procedures can be classified according to the kind of analysis (qualitative, semiquantitative, or quantitative), type of assay format (manual or automated) and assay system (liquid phase, solid–liquid phase, [non-]equilibrium), making use of (radioisotopic or nonisotopic) labeled markers (to detect the antigen–antibody complex) or nonlabeled markers (in which the antigen–antibody complex is detected without labeled markers). The term immunoassay refers to competitive methods while immunometric assays refer to noncompetitive, sandwich-type assay formats. As early as 1969 the first generation of binding assays emerged with development of a radioimmunoassay (RIA) for quantification of insulin antibody formation. Later, the evolution of technical developments led to nonisotopic labels (enzyme-, fluorescence, time-resolved fluoro-, (chemo-) luminescence immunoassays, etc.), monoclonal antibodies, phase matrices, and two-site immunometric sandwich-type assay formats employing two or more antibodies that will bind the analyte at repetitive or different epitope binding sites. The enzyme-linked immunosorbent assay (ELISA) format is

a commonly used type of two-site sandwich type assay in clinical routine work. The analyte is allowed to react noncompetitively with an excess of immobilized (“capture”) antibody (coupled to a solid phase) and after addition and washing off the excess amount of sample specimen, an excess amount of marker labeled (“signal”) antibody is added to bind to another epitope of the analyte. The sandwich thus formed is provided with marker label proportional to the amount of analyte present in the sample. These quantitative assay formats have several advantages: large numbers of specimens can be processed in parallel, providing reproducible results with reasonable precision, sensitivity, specificity, and accuracy. A major advantage of immunometric assay formats over semiquantitative or qualitative techniques is provided by the quantitative endpoint as measured against a defined standard, although there are limitations. Often, the analyte standard is not well defined, the assay procedure is not fully validated prior to use in patient studies, or the possibility to make comparisons to a reference method is lacking. Strict measures of quality assurance or good manufacturing practice protocols are needed to ensure proper assay performance before they should be applied.

Because early detection of small breast lesions is becoming common practice, there is an increasing demand to measure correctly biomarkers in smaller pieces of tumor specimens obtained through fine needle aspiration, core biopsies, or cryostat sections. This implies that there is an immediate need for more sensitive techniques than the standard immunoassays available to date. Alternative approaches to ELISA are proteomic methods such as MALDI and SELDI TOF mass spectrometry (MS), tandem MS, plasma resonance techniques, and antibody chip technologies. Of course the same rigorous principles of quality assurance should be applied for these new methodologies as for the more conventional immunoassays.

3. VARIABILITY IN TEST RESULTS

Assay results are often heterogeneous because of variations in specimen composition, tissue processing, design and specificity of the employed assay, as well as the statistics used for analyzing the collected data. In each of these stages, intrinsic differences in molecular forms (isoforms) of the biomarker present in the tumor tissue are augmented by external causes. The sampling procedure (e.g., fine-needle aspirate, core biopsy, or large biopsy obtained during surgery), the source of tissue (fresh or frozen), storage conditions (time, temperature, freeze–thawing cycles, etc.), and tissue processing (cytosol fraction, membrane extracts) may severely influence the final assay results (3). Likewise, this also holds true for the quantification of biological markers in serum or plasma (4).

Variable design of immuno(metric) assays results in the generation of different test results because different kits incorporate a broad spectrum of

antibodies, sometimes with different antibody specificities and/or affinities. Also, the use of different standards and reference materials provided with the kits are a source of variations in test results. Furthermore, different data reduction processes and statistical techniques are used to analyze tumor marker data and this may lead to a variety of conclusions regarding the clinical interpretation. The computational data processing of laboratory results must be appropriate, uniform, and evaluated extensively (5). McGuire and Clark stated that the design of confirmatory clinical studies should be identical to that of the definite study (2). It is of most importance to note that this also applies to all laboratory steps including tissue storage and processing, the analytical procedures, and the subsequent data processing.

The number and diversity of biomarkers for assessment of cancer prognosis is expanding rapidly, as is the variety of analytical formats and procedures used for quantification. A substantial proportion of assays is based on immunochemical principles and there is a widespread use of nonvalidated assay formats in clinical research settings. Because many assays are poorly standardized and (external) quality control is lacking in most cases, nonvalidated assay results without provided certified guidelines for interpretation become available at a too preliminary stage of assay development. Thus, biomarker testing procedures in laboratories participating in clinical trials should be standardized and externally quality assessed. This requires settlement of quality standards of all assay reagents included in assay kits, provision of guidelines for standardized assay protocols, standardized algorithms for calculation of assay results, and statistical procedures to allow unequivocal interpretation of clinical effect measures. Finally, to ascertain continuity of reliable biomarker data generation, there is a need for guidelines toward uniform internal and external quality assessment procedures. The next sections will discuss preanalytical, analytical, and postanalytical aspects of assay performance.

4. ASPECTS OF BIOMARKER ASSESSMENT

4.1. Preanalytical Criteria

4.1.1. SAMPLING BIAS OF TISSUE SPECIMENS AND TISSUE PROCESSING

Because many tumors are heterogeneous the size of a tumor tissue specimen is important to avoid sampling bias. This bias may lead to different assay results if different areas of a tumor are analyzed (different content of tumor cells, nonmalignant cells, extracellular matrix, fat, and necrotic spots). Thus, fine needle biopsy results may differ from those obtained from a tumor tissue biopsy specimen. Selection bias may occur if frozen tissue specimens from large tumor banks are used in retrospective studies as generally in tumor banks relatively larger samples of frozen tumor tissues are overrepresented (6).