Evaluation of Biomarkers and Surrogate Endpoints in Chronic Disease

ISBN
978-0-309-15129-0

336 pages
6 x 9
PAPERBACK (2010)

Christine M. Micheel and John R. Ball, Editors; Committee on Qualifications of Biomarkers and Surrogate Endpoints in Chronic Disease; Institute of Medicine

EVALUATION OF BIOMARKERS AND SURROGATE ENDPOINTS IN CHRONIC DISEASE

Committee on Qualification of Biomarkers and Surrogate Endpoints in Chronic Disease

Board on Health Care Services

Board on Health Sciences Policy

Food and Nutrition Board

Christine M. Micheel and John R. Ball, *Editors*

INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, N.W. Washington, DC 20001

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This study was supported by Contract No. HHSF223200810020I between the National Academy of Sciences and the Food and Drug Administration. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the organizations or agencies that provided support for this project.

Library of Congress Cataloging-in-Publication Data

Evaluation of biomarkers and surrogate endpoints in chronic disease / Committee on Qualification of Biomarkers and Surrogate Endpoints in Chronic Disease, Board on Health Care Services, Board on Health Sciences Policy, Food and Nutrition Board, Institute of Medicine ; Christine M. Micheel and John R. Ball, editors.
p. ; cm.
Includes bibliographical references.
ISBN 978-0-309-15129-0 (pbk.) — ISBN 978-0-309-15130-6 (pdf) 1. Biochemical markers—Evaluation. 2. Chronic diseases. I. Micheel, Christine. II. Ball, John, 1944- III. Institute of Medicine (U.S.). Committee on Qualification of Biomarkers and Surrogate Endpoints in Chronic Disease.
[DNLM: 1. Biological Markers. 2. Chronic Disease. 3. Evidence-Based Practice. QW 541 E917 2010]
R853.B54E93 2010
610.28—dc22
2010020157

Additional copies of this report are available from The National Academies Press, 500 Fifth Street, N.W., Lockbox 285, Washington, DC 20055; (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area); Internet, http://www.nap.edu.

For more information about the Institute of Medicine, visit the IOM home page at: **www.iom.edu.**

Printed in the United States of America

The serpent has been a symbol of long life, healing, and knowledge among almost all cultures and religions since the beginning of recorded history. The serpent adopted as a logotype by the Institute of Medicine is a relief carving from ancient Greece, now held by the Staatliche Museen in Berlin.

Suggested citation: IOM (Institute of Medicine). 2010. *Evaluation of biomarkers and surrogate endpoints in chronic disease.* Washington, DC: The National Academies Press.

"Knowing is not enough; we must apply.
Willing is not enough; we must do."
—Goethe

INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES

Advising the Nation. Improving Health.

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Charles M. Vest is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. Charles M. Vest are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

COMMITTEE ON QUALIFICATION OF BIOMARKERS AND SURROGATE ENDPOINTS IN CHRONIC DISEASE

JOHN R. BALL (*Chair*), Executive Vice President, American Society for Clinical Pathology
MICHELLE A. ALBERT, Assistant Professor of Medicine, Associate Physician, and Director of Behavioral and Neurocardiovascular Cardiology, Brigham and Women's Hospital, Harvard Medical School
FRED APPLE, Medical Director, Clinical Laboratories, Hennepin County Medical Center, and Professor, Laboratory Medicine and Pathology, University of Minnesota School of Medicine
ROBERT M. CALIFF, Vice Chancellor for Clinical Research and Professor of Medicine, Cardiology, Duke University School of Medicine
VICTOR G. DE GRUTTOLA, Professor and Chair, Biostatistics, Harvard School of Public Health
DAVID L. DEMETS, Professor and Chair, Biostatistics & Medical Informatics, University of Wisconsin–Madison
ROBERT GERSZTEN, Research Director and Faculty Member, Massachusetts General Hospital, and Associate Professor of Medicine, Harvard Medical School
WILLIAM R. HARLAN, JR., Consultant
ALLAN S. JAFFE, Professor of Medicine, Mayo Clinic
RONALD M. KRAUSS, Director, Atherosclerosis Research and Senior Scientist, Children's Hospital Oakland Research Institute
HARLAN M. KRUMHOLZ, Harold H. Hines, Jr., Professor of Medicine and Epidemiology and Public Health, Yale University School of Medicine
MARIA LOPES-VIRELLA, Professor, Bioengineering, Medical University of South Carolina
ROBERTA B. NESS, Dean, University of Texas Health Science Center, School of Public Health
JENNIFER VAN EYK, Associate Professor, Physiology and Biochemistry, Johns Hopkins University
JOHN A. WAGNER, Vice President, Clinical Pharmacology, Merck and Company, Inc.

Consultant

ELIZABETH A. YETLEY, Consultant, National Institutes of Health, Office of Dietary Supplements

Study Staff

CHRISTINE M. MICHEEL, Study Director
SHARYL NASS, Senior Program Officer
ERIN BALOGH, Research Associate
BERNADETTE McFADDEN, Research Associate (from July 2009 to January 2010)
LISA BOYETTE, Christine Mirzayan Science and Technology Policy Graduate Fellow (September to November 2009)
ANNA WOLOSZYNSKA-READ, Christine Mirzayan Science and Technology Policy Graduate Fellow (January to April 2009)
CAIRA M. WOODS, Christine Mirzayan Science and Technology Policy Graduate Fellow (January to April 2010)
DESH MOHAN, Intern (June to August 2009)
ASHLEY McWILLIAMS, Senior Program Assistant
PATRICK BURKE, Financial Associate
ROGER HERDMAN, Director, Board on Health Care Services
LINDA D. MEYERS, Director, Food and Nutrition Board
ANDREW POPE, Director, Board on Health Sciences Policy

Reviewers

This report has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the National Research Council's Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making its published report as sound as possible and to ensure that the report meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process. We wish to thank the following individuals for their review of this report:

RUSS B. ALTMAN, Stanford University
DIANNE BIRT, Iowa State University
JAMES DE LEMOS, University of Texas Southwestern Medical Center at Dallas
THOMAS R. FLEMING, University of Washington
PHILIP GREENLAND, Northwestern University
CHARLES HENNEKENS, Florida Atlantic University
JANE HENNEY, University of Cincinatti
WOLFGANG KOENIG, University of Ulm Medical Center
VICTOR MONTORI, Mayo Clinic
DAVID RANSOHOFF, University of North Carolina at Chapel Hill
CHRISTINE SEIDMAN, Harvard Medical School and Howard Hughes Medical Institute

ALAN TALL, Columbia University
STEPHEN WILLIAMS, Somalogic, Inc.
ALAN WU, University of California, San Francisco

Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations nor did they see the final draft of the report before its release. The review of this report was overseen by **CHARLES C. J. CARPENTER,** the Miriam Hospital, and **KRISTINE M. GEBBIE,** Hunter College, City University of New York. Appointed by the National Research Council and the Institute of Medicine, they were responsible for making certain that an independent examination of this report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this report rests entirely with the authoring committee and the institution.

Preface

Over breakfast during the second meeting of this committee, the members informally discussed a message on the package of one of the cereal offerings, a box of Cheerios. Against the backdrop of an image of a heart, the message was, "You Can Lower Your Cholesterol 4% in 6 weeks." A month later (purely coincidentally), the Food and Drug Administration (FDA) sent a letter to the chair of General Mills, the producer of Cheerios. That letter stated, "based on claims made on your product's label, we have determined that your Cheerios® Toasted Whole Grain Oat Cereal is promoted for conditions that cause it to be a drug because the product is intended for use in the prevention, mitigation, and treatment of disease." Five months later, the new FDA Commissioner Margaret Hamburg indicated, in an Industry Letter, that the agency was examining "Front of Package" (FOP) labels for false or misleading claims, citing consumer studies that found that, with FOP labeling, people are less likely to check the Nutrition Facts Panel, generally found on the side or back of food packages. Notably, H.R. 1105, the Omnibus Appropriations Act of 2009, included funds for an Institute of Medicine (IOM) study to examine and make recommendations regarding Front of Package nutrition symbols.

In the context of the committee's task, this instance illustrates two issues with which the committee wrestled. The first is how science may inform policy decisions when diverse, and sometimes disparate, interests are involved. In this case, consumers wish to choose healthier diets, the food industry has an interest to market its products as healthy, and the FDA needs to minimize risks to the food supply and to inform consumers appropriately. The second is how to make policy decisions before the full process of reaching scientific consensus has been completed.

This report was initiated by the Center for Food Safety and Applied Nutrition of the FDA, which has received dozens of applications for approval of health claims for foods, most of which reflected claims of effects on a biomarker—a patient characteristic that can be measured and is believed to have a significant biological effect. The principal task requested of the Institute of Medicine was to recommend a framework for the evaluation of biomarkers; additionally, the IOM was to make ancillary recommendations for their application. As shown in Chapter 1, however, the task goes beyond claims on foods alone. A framework has been proposed that can be applied across many of the product areas regulated by the FDA.

The Institute of Medicine convened a committee of experts from a variety of related fields, supported by a highly capable technical staff. The committee met face to face four times and had several teleconferences. The committee was further supported by presentations from outside experts in a workshop format, and it benefited from comments from interested parties. As always, the committee's report underwent a rigorous external review, which helped significantly to focus and clarify the findings and recommendations.

The committee met its principal task by recommending a three-part framework for biomarker evaluation: (1) Analytical validation—in essence, is the biomarker able to be accurately measured? (2) Qualification—is the biomarker associated with the clinical endpoint of concern? and (3) Utilization—what is the specific context of the proposed use? The committee met the additional task by making recommendations for implementing the evaluation framework, for supporting evidence-based decision making, and for promoting the public health.

The committee notes that endpoints can be conceptualized in a spectrum. At the end defined by endpoints with less relationship to patient or consumer experience are those that depend on biomarkers alone; in the middle are clinical events that depend on biomarkers as part of the definition; more closely related endpoints are those events that affect patients' lives; and at the near end of the spectrum are the clearest clinical endpoints, such as death. Furthermore, the committee emphasizes that biomarkers cannot be qualified for a use without understanding the specific use and its context.

The committee heard significant evidence of the public's (and professionals') innumeracy, or numerical illiteracy, and the barrier that innumeracy poses to understanding the balance of risk and benefit. Thus, the committee recognizes that significant efforts may be needed, both by government and by professional societies, to inform and educate the public and professionals on how to interpret scientific information so that good science can inform individual decision making.

Critical to the committee's recommendations, and flowing from our consideration of the evidence and vigorous debate, is that there is neither rationale nor scientific basis for predicating regulatory decisions on different levels of scientific evidence for different substances: "science is science." That is, the same level of scientific evidence of benefit and risk should be required of foods as of drugs (and, indeed, of the other substances the FDA regulates—biologics, devices, and cosmetics). The counterargument that some substances (e.g., drugs) pose greater risks than others (e.g., foods) is not dispositive. Counter to that argument is that foods are encountered by a greater population than the target group who encounter drugs, and though drugs are subject to professional mediation (e.g., prescription and counseling), foods are not. As for risk, no one who is allergic to peanuts, eggs, or shellfish would argue that foods are less risky than drugs.

At the risk of using a personal anecdote, I have suffered three episodes of cardiac arrhythmia atrial fibrillation, all associated with drinking two glasses of red wine. Since making the correlation, I've ceased drinking red wine, and have ceased having episodes of atrial fibrillation. When I explained to my elderly mother why I no longer drank red wine, she said, "But I thought red wine was good for you." The answer, of course, is "It depends." "It depends" means that the context of health claims matters. Biomarkers can enable faster, more efficient clinical trials. They can help public health professionals identify and track disease outbreaks. In addition, they can help healthcare practitioners and patients make decisions about care. But the context of their use matters, and the scientific base for their use should be rigorous.

As chair of the committee, I thank personally all the committee members for their individual and group contributions, their diligence, and their comity. I am very grateful for the time and effort that such busy people were willing, often with short turnaround times, to devote to the work of the committee. As is the case with the best of these deliberations, their engaged back-and-forth nature led to a richer, more accurate, and—we all hope—helpful report for regulators, professionals, and the public. None of this could have been accomplished without the professional IOM staff, led by Christine Micheel, who, in addition to her technical expertise, was uncommonly responsive to the committee's direction and to individual comments of the members. I know the committee would join in giving my heartfelt thanks to her.

John Ball, *Chair*
Committee on Qualification of Biomarkers and
Surrogate Endpoints in Chronic Disease

Acknowledgments

The committee and IOM staff would like to thank many individuals for their contributions to this study. Elizabeth Yetley, consultant to the committee, provided needed guidance. We thank Thomas H. Lee, Susan Mayne, Gil Omenn, David Ransohoff, and John Wagner for their project initiation assistance. We thank Joseph Bonventre, Kathleen Ellwood, Federico Goodsaid, and Paula Trumbo for presentations at the first committee meeting. We thank Nancy Cook, Charles Hennekens, and Marshall Joffe for their assistance with editing report sections. We thank all of the workshop speakers for their participation (please see Appendix E for the list of speakers). IOM staff Sharyl Nass, Christine L. Taylor, Roger Herdman, Linda Meyers, and Andrew Pope provided needed assistance. Finally, we thank the FDA for study funding.

The committee would like to thank IOM staff for their assistance with report drafting.

We would like to thank the fellows and interns involved with this study for their assistance. Lisa Boyette contributed to writing and editing tasks. Anna Woloszynska-Read contributed to workshop planning and background research, Caira Woods contributed to report review and creation of report dissemination material, and Desh Mohan provided research and meeting assistance.

Finally, the study director would like to thank project staff for their contributions. Research associates Erin Balogh and Bernadette McFadden were involved in writing many sections of the report. Ashley McWilliams arranged meetings and many other administrative tasks and contributed to research and writing tasks.

Contents

Boxes, Figures, and Tables

Chapter 2

Boxes

Figures

Tables

Chapter 3

Boxes

Appendix A

Tables

Appendix B

Boxes

Summary

Biomarkers are characteristics that are objectively measured and evaluated as indicators of normal biological processes, pathogenic processes, or pharmacologic responses to an intervention. Cholesterol and blood sugar levels are biomarkers, as are blood pressure, enzyme levels, measurements of tumor size from magnetic resonance imaging (MRI) or computed tomography (CT), and the biochemical and genetic variations observed in age-related macular degeneration. Biomarkers can enable faster, more efficient clinical trials for life-saving and health-promoting interventions. They can help improve understanding of healthy dietary choices, and they can help public health professionals to identify and track health concerns. Biomarkers help health care practitioners and their patients make decisions about patient care. The use of biomarkers depends on the quality of data that supports their use and on the context in which they are applied. Evaluation of the quality of the measurements and data linking the biomarkers to clinical outcomes is important for assessing biomarker utility.

The Food and Drug Administration (FDA) requested the Institute of Medicine (IOM) to recommend a framework for the evaluation of biomarkers. The committee has recommended such a framework, with critical components of analytical validity, evidentiary qualification, and utilization analysis (Box S-1). The framework is intended to bring consistency and transparency to a previously non-uniform process. During its deliberations, the committee identified a need for the FDA to evaluate biomarker use with the same degree of scientific rigor across the product categories regulated by the agency, including drugs, biologics, devices, foods, and supplements. The committee has also recommended strategies for implementing the evaluation framework, supporting the use of evidence-based regulation and the protection and promotion of public health.

BOX S-1
Summary of Recommendations for Effective Biomarker Evaluation

The Evaluation Framework

1. The biomarker evaluation process should consist of the following three steps:

1a. Analytical validation: analyses of available evidence on the analytical performance of an assay;

1b. Qualification: assessment of available evidence on associations between the biomarker and disease states, including data showing effects of interventions on both the biomarker and clinical outcomes; and

1c. Utilization: contextual analysis based on the specific use proposed and the applicability of available evidence to this use. This includes a determination of whether the validation and qualification conducted provide sufficient support for the use proposed.

2a. For biomarkers with regulatory impact, the Food and Drug Administration (FDA) should convene expert panels to evaluate biomarkers and biomarker tests.

2b. Initial evaluation of analytical validation and qualification should be conducted separately from a particular context of use.

2c. The expert panels should reevaluate analytical validation, qualification, and utilization on a continual and a case-by-case basis.

Scientific Process Harmonization

3. The FDA should use the same degree of scientific rigor for evaluation of biomarkers across regulatory areas, whether they are proposed for use in the arenas of drugs, medical devices, biologics, or foods and dietary supplements. Congress may need to strengthen FDA authority to accomplish this goal.

4. The FDA should take into account a nutrient's or food's source as well as any modifying effects of the food or supplement that serves as the delivery vehicle and the dietary patterns associated with consumption of the nutrient or food when reviewing health-related label claims and the safety of food and supplements. Congress may need to strengthen FDA authority to accomplish this goal.

Biomarkers are measurements that indicate biological processes (see Box S-2 for definitions of key terms). Biomarkers include physiological measurements, blood tests, and other chemical analyses of tissue or bodily fluids, genetic or metabolic data, and measurements from images. Cholesterol and blood sugar levels are biomarkers, as are blood pressure, enzyme levels, measurements of tumor size from MRI or CT, and the biochemical and genetic variations observed in age-related macular degeneration. Emerging technologies have also enabled the use of simul-

BOX S-2
Important Definitions

Analytical Validation: "assessing [an] assay and its measurement performance characteristics, determining the range of conditions under which the assay will give reproducible and accurate data."[a]

Biomarker: "a characteristic that is objectively[b] measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a[n] . . . intervention."[c] Example: cholesterol level.

Chronic Disease: a culmination of a series of pathogenic processes in response to internal or external stimuli over time that results in a clinical diagnosis/ailment and health outcomes. Example: diabetes.

Clinical Endpoint: "a characteristic or variable that reflects how a patient [or consumer] feels, functions, or survives."[c] Example: death.

Fit-for-Purpose: being guided by the principle that an evaluation process is tailored to the degree of certainty required for the use proposed.

Qualification: "evidentiary process of linking a biomarker with biological processes and clinical endpoints."[d]

Surrogate Endpoint: "a biomarker that is intended to substitute for a clinical endpoint. A surrogate endpoint is expected to predict clinical benefit (or harm or lack of benefit or harm) based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence."[c] Example: blood pressure for trials of several classes of antihypertensive drugs.[e]

NOTES: [b] The committee defines "objectively" to mean "reliably and accurately." [e] Please see Chapter 2 for discussion of this biomarker.
SOURCES: [a] Wagner (2002); [c] Biomarkers Definitions Working Group (2001); and [d] Wagner (2008).

taneously measured "signatures," or patterns of co-occurring sets, of genetic sequences, peptides, proteins, or metabolites as biomarkers. These signatures can also be combinations of several of these types of measurements; ideally, each component of a signature is identified.

Biomarkers are used to describe risk, exposures, intermediate effects of treatment, and biologic mechanisms; as surrogate endpoints, biomarkers are used to predict health outcomes. Biomarkers can provide information about risk and physiological parameters that is useful in a variety of contexts: (1) insight into the health and well-being of patients and consumers, (2) the status of patient and consumer response to an intervention, (3) a basis for interpreting research results and comparing results across studies, (4) indications of health status and disease risk in population groups, and (5) important data for planning and evaluating public health programs. Biomarker measurements support the practice

of modern medicine; the development of effective drugs, biologics, and devices; the communication of information about healthy food[1] choices and dietary habits; and the planning and monitoring of public health initiatives; in some circumstances, use of biomarkers is essential for these goals. A variety of biomarkers and uses have advantages for patients and consumers, physicians and other healthcare practitioners, scientists and researchers, industry, payers, regulators, and policy makers.

It is important to note the distinction between biomarkers, risk factors, and endpoints. Biomarkers are patient and consumer characteristics that are measured and evaluated. As measurements, they are subject to measurement quality issues such as accuracy, precision, reliability, reproducibility, and the need for standards and quality control. Risk factors are variables that predict outcomes and are composed of biomarkers and social and environmental factors. The value of a risk factor depends on the degree to which it can predict an event. Finally, there are endpoints—which often include biomarkers, alone or in combination with clinical events. Endpoints range from something a patient or consumer clearly experiences, such as mortality, or a variable that is to some degree related to events impacting a patient or consumer's life. An example of an endpoint more closely related to patient or consumer experience would be acute myocardial infarction with full recovery and without impact on a patient or consumer's quality of life, and a less clearly related example is an LDL cholesterol level (more accurately, non-HDL cholesterol), as associated with cardiovascular disease mortality. The value of an endpoint increases in relation to the degree to which it conveys information about the effect of an intervention on a patient or consumer's experience of life. For endpoints that are less clearly related to patient or consumer experience, there is a need to acknowledge that we cannot know with certainty whether a beneficial change in the endpoint will impact a patient or consumer's experience of life. Further, the committee notes that endpoints can be conceptualized in a spectrum. At one end are endpoints defined by biomarkers alone that have less relationship to patient or consumer experience; in the middle are clinical events that depend on biomarkers as part of the definition; further along the spectrum are endpoints that are more closely related to events that affect patients' and consumers' lives; and at the other end of the spectrum are the clearest clinical endpoints, such as death.

[1] In this report, the term *food* is inclusive of foods consumed as part of meals and snacks, dietary supplements, and components contained in them (nutrients, other bioactive substances).

STUDY SCOPE

Following the recommendations from the 2007 Institute of Medicine report *Cancer Biomarkers: Challenges of Improving Detection and Treatment* (IOM, 2007), the Center for Food Safety and Applied Nutrition of the FDA asked the IOM to generate recommendations on the evaluation process for biomarkers, with focus on biomarkers and surrogate endpoints in chronic disease. The committee was to recommend a framework for biomarker evaluation and test it using case studies of biomarkers and surrogate endpoints in various diseases, including low-density and high-density lipoprotein cholesterol levels as biomarkers of coronary heart disease.

Focusing on this charge, the committee outlined considerations for determining the appropriate use of biomarkers across a variety of contexts, including foods, drugs, biologics, and devices.

FINDINGS, CONCLUSIONS, AND RECOMMENDATIONS

The recommendations developed by the committee fall into two main categories: the biomarker evaluation process and strengthening evidence-based regulation. Recommendation 1 is meant to be applicable to all uses of biomarkers. Recommendations 2, 3, and 4 are focused on uses of biomarkers that result in regulatory decisions and the impacts these decisions have on public health, whether for drugs, biologics, or device development; for relationships between diet or nutrients/food substances and disease; or for public health monitoring and interventions. Recommendations 5 and 6 are ancillary recommendations that provide for efficient and effective implementation of Recommendations 1–4. The report will explain why scientific rigor is important when describing relationships among food, biomarkers, and chronic disease. This report uses biomarkers of cardiovascular disease for many of its illustrative examples, but examples from other diseases are also considered.

Biomarker Evaluation Process

The committee concluded that it was important to address several challenges revealed by previous biomarker evaluation efforts. First, preanalytical and analytical validation of biomarker tests has often been underemphasized in that it has not been considered an integral component of biomarker qualification. Therefore, the committee has included preanalytical and analytical validation as a necessary component, and it has used the term "biomarker evaluation" to include both validation and qualification. Second, in general, the evidentiary assessment and utilization or context-of-use components of qualification are not adequately separated. The committee's proposed process separates these steps so that

the different investigative and analytical processes required to evaluate evidence and contexts of use are defined. Finally, previous evaluation frameworks have not explicitly incorporated a process for reevaluation of analytical validation, evidentiary assessment, and context of use based on new data. The committee also recognizes that some biomarker evaluation steps may occur concurrently.

The evaluation framework is intended to be applicable across a wide range of biomarker uses, from exploratory uses for which less evidence is required to surrogate endpoint uses for which strong evidence is required. The framework is meant for, but not limited to, use in research, clinical, product, and claim development in food, drug, and device industries, and public health settings, and it is intended to function for panels of biomarkers in addition to single biomarkers and for circulating, genetic, and imaging biomarkers. The committee employed case studies to illustrate the use of the evaluation framework because different biomarkers and uses will emphasize different aspects of the general principles set forth in the report.

Recommendation 1:
The biomarker evaluation process should consist of the following three steps:

1a. Analytical validation: analyses of available evidence on the analytical performance of an assay;

1b. Qualification: assessment of available evidence on associations between the biomarker and disease states, including data showing effects of interventions on both the biomarker and clinical outcomes; and

1c. Utilization: contextual analysis based on the specific use proposed and the applicability of available evidence to this use. This includes a determination of whether the analytical validation and qualification conducted provide sufficient support for the use proposed.

It is important to emphasize that the steps listed above are interrelated; they are not necessarily separated in time, and conclusions in one step may require revisions or additional work in other steps (see Figure S-1).

Recommendation 2 provides further guidance on the application of the framework to uses of biomarkers that have regulatory impact. Specifically omitted from this recommendation are biomarker discovery activities and biomarkers for use in drug discovery, development, or other preclinical uses. The committee sought ways to achieve a rigorous evaluation framework without stifling innovation. Experts qualified by experience

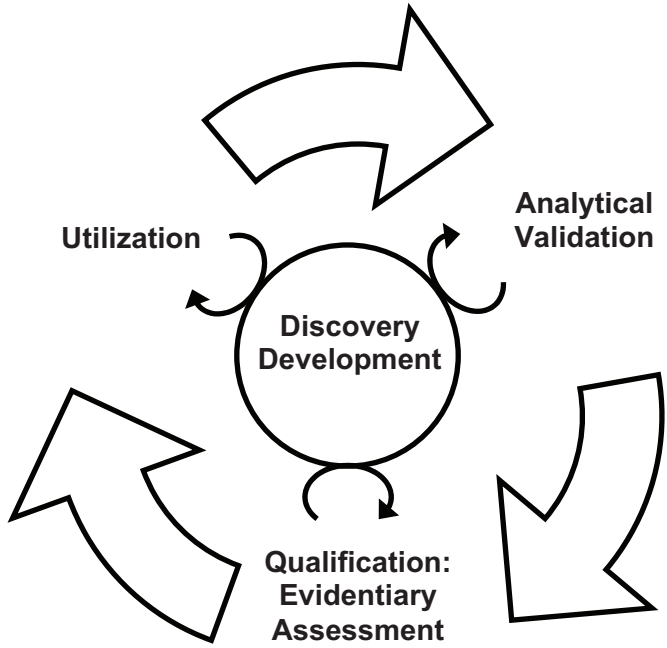

FIGURE S-1 The steps of the evaluation framework are interdependent. While a validated test is required before qualification and utilization can be completed, biomarker uses inform test development, and the evidence suggests possible biomarker uses. In addition, the circle in the center signifies ongoing processes that should continually inform each step in the biomarker evaluation process.

and training are needed to conduct the evaluation reviews, focusing on the utilization step, because case-by-case analyses are the only way to ensure proper use of biomarkers given the state of the science.

Due to the complexity and progressive increase in the amount of data, the need for fit-for-purpose and context-of-use analysis, and the need to deal with sometimes contradictory evidence, expert input is essential to provide scientific judgment in areas of uncertainty. Likewise, as evidence evolves even after a biomarker is evaluated, it is imperative that biomarkers be reevaluated on a continuing basis so that both the scientific evidence and context-of-use analyses capture the current state of the science. Recommendation 2 will be discussed in the context of each of the three steps of Recommendation 1.

Recommendation 2:

2a. For biomarkers with regulatory impact, the FDA should convene expert panels to evaluate biomarkers and biomarker tests.

2b. Initial evaluation of analytical validation and qualification should be conducted separately from a particular context of use.

2c. The expert panels should reevaluate analytical validation, qualification, and utilization on a continual and a case-by-case basis.

Biomarker evaluation is a dynamic process. By considering additional evidence, it is possible that the expert panel may alter its past findings by revoking recommendations for a previously accepted biomarker use, choosing not to recommend a biomarker for uses similar to those for which it was granted permission in the past, providing a more nuanced explanation as to how a biomarker should be used, or qualifying the biomarker for use in new contexts. The panels may resemble FDA advisory committees. The panelists should possess relevant scientific expertise and experience; a variety of stakeholders should have opportunity for input; and attention should be paid to conflict-of-interest standards in a manner similar to government and IOM advisory committees. By continual, the committee refers to the need for regular reevaluation on the basis of new scientific developments and data.

Analytical Validation

The first step of the proposed evaluation framework is to catalogue the data addressing the analytical validity of the biomarker in question. In the utilization step of the framework, evaluators will determine whether a suitable biomarker test possesses appropriate validation given the proposed use of the biomarker or whether further data gathering is needed. As mentioned earlier, preanalytical and analytical validation is a necessary prerequisite for biomarker qualification. The terminology used in the recommendation, analytical performance, is not meant to describe how well a biomarker correlates with the clinical outcomes of interest. Instead, analytical validation of an assay includes the biomarker's limit of detection, limit of quantitation, reference (normal) value cutoff concentration, and the total imprecision at the cutoff concentration. Depending on the use, biomarker tests need to be reliable, need to be reproducible across multiple laboratories and clinical settings, and possess adequate sensitivity and specificity for the biomarker being measured before data based on their use can be relevant in the subsequent biomarker evaluation steps. Appropriate standards for ensuring quality and reproducibility in different clinical and laboratory settings and across relevant populations should be available. Validation of biomarker tests should be done on a test-by-test

basis and must then be deemed sufficient for the use proposed in the utilization step. Validation may also include efforts to determine the extent for which data from different tests for the same biomarker may be compared to one another. When comparability is achieved, it both strengthens the biomarker itself and adds power to retrospective analyses of data related to the biomarker. As indicated in Recommendation 2, the expert panel will need to reevaluate the validation assessments on a continuing and as-needed basis and evaluate new tests that become available.

Qualification

The second step of the committee's evaluation framework incorporates a factual description of the available evidence. The first component of qualification is to evaluate the prognostic value of the biomarker–disease relationship, or the nature and strength of evidence about whether the biomarker is associated with disease outcomes. This is discussed further below. The second component is to gather available evidence showing the biomarker's ability to predict the effects of interventions on clinical endpoints of interest; this evidence may also be used to support the associations described in the first component. If the biomarker–clinical endpoint relationship persists over multiple interventions, it is considered more generalizable. It is important to note, however, that the type of reasoning that may be used in qualification is probabilistic rather than deterministic. Although deterministic reasoning ultimately means that every contributing factor to the biomarker–intervention–clinical endpoint link is defined and understood, probabilistic reasoning emphasizes epidemiological and statistical relationships, acknowledging that all contributing factors are generally not fully understood and that some factors may be fundamentally random.

Related to the first component of qualification, prognostic value can be assessed by using concepts described by criteria proposed for establishing causation of non-infectious diseases (Advisory Committee to the Surgeon General, 1964; Hill, 1965). These criteria evaluate characteristics such as temporality, strength of association, biological plausibility, and consistency, among others. Given that biomarkers are "indicators"—in that they are not necessarily causal—and that an abnormal value or a gradient in level over time is not necessarily informative or predictive depending on the clinical situation, the committee instead used these criteria as a structure for assessing the prognostic value, or degree of association between the biomarker and the clinical outcomes of interest absent any interventions. For a surrogate endpoint, or a biomarker deemed useful as a substitute for a defined, disease-relevant clinical endpoint, prognostic value is a necessary—but not sufficient—criterion for

the evaluation. Depending on the situation, not all of the criteria must be fulfilled; temporality, strength of association, and consistency are particularly important, however. Observational data in human populations and preliminary clinical data (e.g., phase I or II data) are considered. Nonetheless, determination of whether a biomarker can be used as a surrogate endpoint for a specified intervention is done in the utilization step of the evaluation process.

To address the second component of qualification, robust, adequately controlled clinical study data using clinical endpoints (i.e., phase III data or equivalent studies) are necessary. In the description of the evidence about the biomarker, applicable populations and conditions for use need to be articulated and taken into consideration in the utilization step of the biomarker evaluation framework for all types of proposed uses, including those for dietary and nutritional purposes.

Utilization

The third step of the committee's biomarker evaluation framework is a contextual analysis of the available evidence about a biomarker with regard to the proposed use of the biomarker. It is most essential that this analysis be carried out by a panel of experts, as scientific and medical judgment is necessary to weigh the possible advantages and disadvantages of the proposed biomarker use. These evaluations should take place on a per use basis, because use depends on the context of use proposed and because knowledge and technology continually evolve. Applicable populations and conditions for use need to be articulated. Utilization can be divided into several components. The first is a determination of the general category of use for which the biomarker is intended (e.g., prevention in the general population or a diseased population, diagnosis, treatment, or mitigation); this can guide the panel in determining important factors to consider in the second component of utilization. The second component is consideration of factors such as the prevalence, morbidity, and mortality of the disease; the risks and benefits associated with the intervention; opportunity cost; and whether the biomarker is being considered for use as a surrogate endpoint.

Strong evidence and a compelling context are needed for the utilization of a biomarker as a surrogate endpoint in situations with regulatory impact. In the case of chronic disease, where there are multiple pathogenetic pathways leading to development of clinical outcomes and multiple manifestations of disease, the probabilistic nature of predictions made using biomarker data means that no biomarker can give absolute certainty of an event's future occurrence nor absolute certainty of the timing of

the predicted event. Nonetheless, there are situations in which use of a biomarker as a surrogate endpoint in situations with regulatory impact may be supported, such as in situations where the need for interventions is urgent or where studies including clinical endpoints are not feasible because of technical or ethical reasons. Situations with regulatory impact are defined in Chapter 3. Again, this is not meant to discourage use of biomarkers in product development; biomarkers play an important role in research and decision-making. Finally, it is essential to remember that the information that an individual surrogate endpoint or clinical endpoint can give is inherently limited; as a result, it is important to emphasize the need to evaluate data relating to adverse events and unintended effects of biomarker use. As will be discussed and shown in Chapters 3 and 4, the status of a biomarker as a surrogate endpoint is context specific, and a biomarker cannot be assumed to be a general surrogate endpoint separate from a designated use.

The committee does not intend to imply that selection of endpoints for clinical trials would be simple or risk free if investigators were simply to avoid surrogate endpoints. Clinical and surrogate endpoints have been defined in a way that may imply a clear distinction between the two, in that clinical endpoints typically reflect patient or consumer experience and surrogate endpoints do not. However, there is discussion surrounding this issue, which illustrates the scientific complexity of the distinction between clinical and surrogate endpoints. Some clinical endpoints have many similarities with biomarkers, and can be thought of as a step removed from patient or consumer experience, and therefore subject to similar potential failings as surrogate endpoints (i.e., pain scales). Some surrogate endpoints are highly robust (e.g., HIV-1 RNA for effectiveness of antiretroviral medications in the treatment of HIV infection). Clinical endpoints share many features of biomarkers, such as the need for analytical validation, but they differ from biomarkers in that clinical endpoints address how a patient or consumer feels, functions, or survives and also commonly utilize multiple diagnostic criteria. The committee recognizes that selection of clinical endpoints is beyond the scope of this report. Nonetheless, there are many important interests at stake in this discussion and some issues, such as the best way to choose endpoints for trials, may be context specific. In such settings, stakeholders such as industry, the public as represented by government and community representatives, and academic researchers may benefit from convening to discuss these issues.

Scientific Process Harmonization

Recommendation 3:
The FDA should use the same degree of scientific rigor for evaluation of biomarkers across regulatory areas, whether they are proposed for use in the arenas of drugs, medical devices, biologics, or foods and dietary supplements.

The importance of rigorous biomarker evaluation has been discussed for decades in the context of drug development. For foods, supplements, and devices, however, based on legislative and legal mandates, the FDA's regulation of claims and the scientific standards for evaluating such claims are governed by different regulatory frameworks as compared to drugs; legislation may be required to revise the science-based standards and regulatory processes for these non-drug products. The committee concluded that the same standards of scientific evidence are required across regulatory areas and different products in the various FDA centers as well as for comparative effectiveness research because decisions about foods, drugs, biologics, and devices need to evaluate the evidence for claimed benefits within the context of use. The public health implications are important, and a critical evaluation of the strength of the evidence on safety is an important component of the context-of-use considerations for health claims on foods. Although it may be tempting to assume, for example, that health claims on foods have less potential risk for adverse consequences than is the case for drugs, it is important to realize that health claims on foods potentially impact a far greater portion of the population than do drug claims, that health claims are not interpreted with the mediation of a trained health professional, and that misleading or poorly substantiated health claims—or those later discovered to be incorrect due to insufficient evidence—can result in harm. These potential harms emphasize the need to weigh a biomarker's potential context of use in the utilization step.

The committee's biomarker evaluation framework is intended to accomplish the goal of consistent evaluation of biomarkers across different types of products and contexts of use. The committee recognizes the differences between scientific assessments of data and policy decisions. The first two steps of the evaluation framework are scientific steps. The third step provides a framework in which scientists and other experts can use rigorous scientific information to make recommendations for complex policy decisions.

Recommendation 4:
The FDA should take into account a nutrient's or food's source as well as any modifying effects of the food or supplement that serves

as the delivery vehicle and the dietary patterns associated with consumption of the nutrient or food when reviewing health-related label claims and the safety of food and supplements.

Drugs, biologics, and devices are evaluated for efficacy and safety on the basis of the whole products. Recommendation 4 seeks to extend this approach to foods and supplements. The differing health effects of individual nutrients or other food substances in food or supplement products composed of multiple substances are important. Due to this, for foods, focusing on a single nutrient or food substance contained in a food or in several different foods can be misleading because it fails to take into account potential modifying effects of the source of the substance and matrix effects of other components in the food, meal, and diet. When these evaluations are taking place based on biomarker data, the difficulties that arise due to incomplete data on unintended effects and side effects are compounded. While review of proposed health claims takes into account the relationship of the specific substance that is the subject of the health claim to the health outcome of interest, it may not adequately consider the modifications of the substance's effect on the disease outcome by other bioactive components in that food or the diet.

An individual substance or product composed of multiple substances may impact one or more biological pathways, each raising or lowering risk for a chronic disease or condition. An intervention may also have multiple health outcomes, and although it would be difficult or infeasible to discover or assess all of these effects, it is important to acknowledge them. Figure S-2 illustrates the multiplicity of possibilities inherent in the presence of multiple ingredients, each potentially impacting multiple pathways, in turn leading to multiple outcomes.

Ancillary Recommendations

Effective implementation of the committee's biomarker evaluation framework process across all contexts of use will benefit from coordination within the FDA and with other government agencies. Useful components of this coordination include the systematic collection of data, building and supporting needed information technology infrastructure, and strengthening the surveillance systems required for linking biomarker and clinical outcome data. The FDA needs these tools to gather and use evidence when making the regulatory decisions, which have important effects across the spectrum of research, clinical practice, and public health surveillance. Recommendations 5 and 6 address this need.

Recommendations 5 and 6 are listed in Box S-3 and are discussed in detail in Chapter 5.

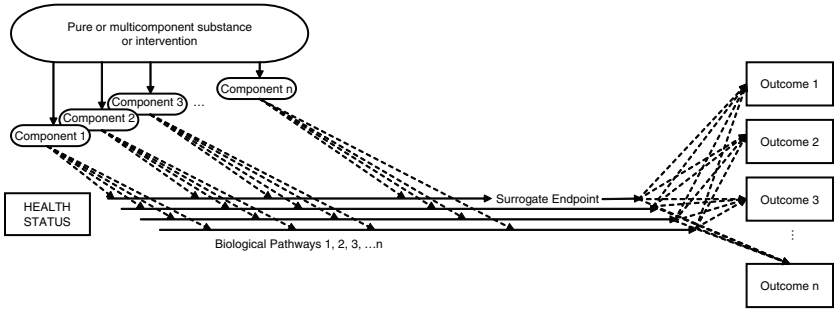

FIGURE S-2 Multiple ingredients, multiple biological pathways, and multiple outcomes illustrate some of the complexities of the use of biomarkers and surrogate endpoints in chronic disease. Note that while the solid horizontal arrows indicate biological pathways, they do not necessarily indicate pathways of the particular disease or condition that a substance or intervention is meant to address. In other words, a surrogate endpoint may not be on the causal pathway of the disease process and a substance or intervention may have mechanisms of action independent of the disease process. Dotted lines indicate possible pathways.

BOX S-3
Ancillary Recommendations

Improving Evidence-Based Regulation

5a. Congress should strengthen the FDA's authority to request and enforce postmarket surveillance across drugs, devices, and biologics when approvals are initially based on putative surrogate endpoint data.

5b. Congress should grant the FDA authority to request studies and sufficient authority to act on the results of studies on consumer understanding of claims on foods and supplements.

6a. The U.S. Department of Health and Human Services (HHS) should facilitate a coordinated, department-wide effort to encourage the collection and sharing of data about biomarkers for all uses, including drugs, biologics, devices, and foods.

6b. The FDA in coordination with other federal agencies should build needed data infrastructure and surveillance systems to handle the information necessary to gain sufficient understanding of the effects of biomarker utilization.

REFERENCES

Advisory Committee to the Surgeon General. 1964. *Report of the Advisory Committee to the Surgeon General.* Washington, DC: U.S. Department of Health, Education, and Welfare.

Biomarkers Definitions Working Group. 2001. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology and Therapeutics* 69(3):89–95.

Hill, A. B. 1965. The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine* 58:295–300.

IOM (Institute of Medicine). 2007. *Cancer biomarkers: The promises and challenges of improving detection and treatment*. Washington, DC: The National Academies Press.

Wagner, J. A. 2002. Overview of biomarkers and surrogate endpoints in drug development. *Disease Markers* 18(2):41–46.

Wagner, J. A. 2008. Strategic approach to fit-for-purpose biomarkers in drug development. *Annual Review of Pharmacology and Toxicology* 48:631–651.

1

Introduction

Biomarkers are tools used by doctors, scientists, and other health professionals to obtain information about a patient's or research subject's health status or response to interventions. Many medical or lifestyle interventions, indispensible to modern medical care, can induce changes in biomarkers. In order for consumers, physicians, drug developers, and policy makers to make informed decisions based on biomarkers, it is important to understand the amount, strength, and quality of data supporting the use of any specific biomarker to direct decisions in clinical care, drug development, public health, and health policy decisions.

Every time a parent takes a child's temperature looking for a fever, they are using a biomarker to assess for illness. That parent may go on to monitor their child's temperature over the course of several days, both to follow the progression of an infection and to determine whether antipyretic and antimicrobial therapies are working effectively. Even this fairly simple example of a biomarker highlights some of the issues associated with their use. For example, the method used to measure body temperature matters. Using a thermometer is a more accurate approach than a hand to the forehead. Slightly different temperatures will be obtained depending on whether the measurement is an oral, ear, rectal, or axillary temperature. Although a fever is a useful piece of information about how a disease process is developing, it is only one piece of information in what could be a complex illness. To further complicate matters, some diseases present with relapsing and remitting fevers, and interpretation of temperature data in that patient population needs to be very different

than an illness where a fever accompanies acute infection and resolution of the fever signals a shift to resolving the infection.

In an ideal setting, biomarkers reflect disease course and activity; many good biomarkers are useful in monitoring disease process and complications. In the diagnosis and management of prostate cancer, for example, prostate-specific antigen (PSA) can be measured in a patient's blood, and PSA levels can be followed as an indicator of whether the cancer is growing or responding to treatment. However, this example illustrates several challenges of using biomarkers. PSA may be elevated in some patients because they have prostate cancer, but it can also be elevated for other reasons. One important finding that has been reported recently is that PSA is not necessarily a good biomarker for population-wide screening for prostate cancer (Sardana et al., 2008). This illustrates the point that biomarkers are effective only to the degree that they are used in the appropriate context. It is critical to note that even a perfect biomarker cannot, with certainty, be used in place of patient outcomes in the evaluation of an intervention.

One step in supporting regulators is to institute an evidence-based, transparent process for biomarker evaluation. Biomarker evaluation is often thought of as two unlinked steps: analytical validation of biomarker tests and biomarker qualification. Biomarker qualification is the evidence-based process of linking a biomarker with one or more clinical endpoints. Decisions to use biomarkers are dependent on the intended applications. Currently, the evaluation of biomarkers is not based on uniform standards or processes, but rather on the gradual development of consensus in the scientific community. The potential value and impact of a more uniform and transparent evaluation process was noted in the 2007 Institute of Medicine (IOM) report, *Cancer Biomarkers: The Promises and Challenges of Improving Detection and Treatment* (IOM, 2007), which recommended that government agencies and non-governmental stakeholders "should work together to develop a transparent process for creating well-defined consensus standards and guidelines for biomarker development, validation, qualification, and use to reduce the uncertainty in the process of development and adoption."

The *Cancer Biomarkers* recommendation gains even more weight when considered with the emergence of pharmacogenetics, pharmacogenomics, and all of the promising medical breakthroughs of personalized medicine. Pharmacogenetics is the science of understanding how an individual's genes may interact to impact drug function and metabolism. Personalized determination of drugs that will work for given patients and dosing based on their metabolic profiles has the potential to decrease unnecessary or not helpful treatments and decrease adverse effects from treatments when they are helpful. Pharmacogenomics is the science of understanding genetic variations between populations in disease incidence, progression,

and treatment. More detailed understanding of disease biology has the potential to lead to more effective prevention and treatment approaches. Biomarkers are critical to progress in these areas, and it will be important that newly discovered biomarkers be adequately studied before being adopted into routine clinical management of patients.

ORIGIN OF THE TASK

In 2008, the Food and Drug Administration's (FDA's) Center for Food Safety and Applied Nutrition (CFSAN), in conjunction with the FDA's Center for Drug Evaluation and Research, approached the IOM for advice on the topic of biomarker and surrogate endpoint evaluation, noting the limited number of surrogate endpoints available, the high cost of evaluating possible surrogate endpoints biomarkers, and the absence of an agreed-upon, systematic, transparent process for biomarker evaluation. Study developers were also interested in learning whether principles of biomarker qualification or evaluation learned in the drug development setting would also be generally applicable in other FDA-regulated product categories, such as foods and supplements. As part of its efforts within the Critical Path Initiative (CPI),[1] CFSAN requested that the IOM charge an expert committee with the following task:

> An Institute of Medicine (IOM) committee will be convened to generate recommendations on the qualification process for biomarkers, with a focus on risk biomarkers and surrogate endpoints in chronic disease. These recommendations will consider existing prototypes for qualification of biomarkers used in drug development. The committee will recommend a framework for qualification and test it using case studies of risk biomarkers and surrogate endpoints for coronary heart disease (CHD) such as low-density lipoprotein (LDL) and high-density lipoprotein (HDL) cholesterol levels. In particular, the committee will:
>
> 1. Conduct a review of current approaches to qualifying biomarkers.
> 2. Recommend a framework that can be used to rank biomarkers according to the types and quality of evidence, considering context of use for a range of product types.
> 3. Demonstrate applications through case studies.
> 4. Make ancillary recommendations for the application, enhanced development, and use of risk biomarkers and surrogate endpoints in chronic disease.[2]

[1] See http://www.fda.gov/oc/initiatives/criticalpath.

[2] The terminology in the statement of task differs in a few ways from the terminology of this report. As will be explained in Chapter 3, the committee's terminology replaces qualification with evaluation in many instances, and risk biomarker with biomarker.

CPI is a framework created by the FDA under which the challenges posed by increasing medical product development costs and lengthening time-to-market for medical products can be addressed. The need for improvement in the process for evaluation of biomarkers and surrogate endpoints was identified from the inception of the CPI at the FDA (FDA, 2004a), and formally recognized as a "Critical Path Opportunity" at CFSAN shortly thereafter (FDA, 2006). The following is an excerpt from the report published in June 2008 describing CFSAN's 2007 progress in this area (FDA, 2008):

> [The] FDA is exploring development of a framework for validating modifiable risk factors (biomarkers) for chronic diseases, such as cancer, heart disease, diabetes, and others that can be the subject of a health claim. The framework will consist of defining the level and type of evidence that is required to support a biomarker that modifies the risk of disease. The first step toward defining a framework will consist of working through the National Academy of Sciences, Institute of Medicine, to convene a panel of experts to outline the steps necessary for qualifying a biomarker for evidence-based decision making, assuming funding becomes available. The task for the panel will be to hold workshops as needed and then to issue a report that [the] FDA can use in its review of scientific evidence offered to substantiate health claims that can be used on food products, including dietary supplements. Funds from the Critical Path [I]nitiative have enabled CFSAN to develop a task order with IOM for this initiative.

Biomarkers and the FDA

With regard to biomarkers, the FDA is subject to competing forces and is expected to evaluate many factors with a limited number of resources. The desire for effective new drugs, devices, and biologics accompanied by the goal of reducing the monetary cost and time expended on development of interventions for chronic diseases serve as incentives for more aggressive use of biomarkers (IOM, 2006). The need to protect patients and consumers from undefined risks is an incentive for more conservative use of efficacy biomarkers and for the development of effective safety biomarkers.

Little consistent, reliable information is currently available regarding how consumers can know which foods might have health benefits beyond basic nutrition. Recently questions have arisen related to use of biomarkers in substantiating health claims about foods, namely whether the use of biomarkers to draw conclusions about the health benefits of nutrients, foods, and supplements should be encouraged, and how information about the uncertainty associated with using biomarkers in this way can be communicated to consumers.

Drug development costs have been estimated at $500 million to $2 billion per product depending on the size of the pharmaceutical company (Adams and Brantner, 2006). CPI began a few years after the implementation of accelerated approval regulations, and it identified a need for more biomarkers of efficacy. Public–private partnerships, such as the Critical Path Institute and the Biomarkers Consortium, were formed, in part, to foster precompetitive data sharing related to biomarker development. The Biomarkers Consortium[3] has brought together industry, academia, the FDA, the National Institutes of Health (NIH), and the Centers for Medicare & Medicaid Services to identify and address areas of greatest potential impact in the need for new qualified biomarkers. However, their focus is primarily on facilitating the discovery of new biomarkers. As a result they have not made it a priority to propose an evaluation framework for biomarkers.

At the start of CPI in 2004, it was estimated that only 8 percent of medicinal compounds reaching phase I clinical trials would eventually be approved for marketing (FDA, 2004b). One of the primary ways that CPI proposed to speed approvals was through the use of biomarkers. With accelerated approval came a greater need for postmarket studies of approved medicinal products. The FDA has faced and attempted to resolve some administrative challenges, such as manufacturers' nondisclosure and/or underreporting of adverse events that result from product usage; inadequate resources to strengthen and broaden oversight efforts; and antiquated information technology systems, in effectively requesting and enforcing these studies, as will be discussed in Chapter 5.

Nutrients, foods, and supplements are regulated under a different framework than are drugs, devices, and biologics. The FDA regulates products purchased with one out of every four consumer dollars spent. Of this amount, 75 percent is spent on products regulated by CFSAN: foods, supplements, and cosmetics. CFSAN's $470 million budget regulates the $525 billion food and cosmetics industry (FDA, 2009a). Foods do not undergo premarket evaluation. New ingredients are evaluated, but for safety only. CFSAN also regulates the labeling of foods. This includes the familiar nutrition facts panel as well as a variety of health-related claims found on food labels and promotional materials.

To a certain extent, the FDA's evaluation of health claims has been crippled by the lack of an agreed-upon, transparent process for biomarker evaluation. Authorized and qualified health claims, which describe links between a food substance and a reduction in risk for a disease, may include data based on the measurement of surrogate endpoints or risk biomarkers as justification for the claims. It is uncommon for produc-

[3] See http://www.biomarkersconsortium.org.

ers of foods or supplements to study the effects of foods and nutrients on clinical endpoints, which makes data from surrogate endpoints and biomarkers the focus of applications for health claims. These include folic acid for reducing the risk for neural tube defects and soluble oat fiber for reducing the risk of heart disease. Claims must be evaluated and authorized by the FDA in most cases. In some cases, health claims can be authorized based on a statement from an authoritative body, such as the NIH or the National Academy of Sciences.[4] The lack of an agreed-upon, transparent process for biomarker evaluation has been seen as one of the roadblocks to a broader selection of surrogate endpoints on which claims could be based.

DEFINITIONS

The committee observed a great deal of inconsistent and imprecise definition and use of terms relevant to biomarkers and biomarker evaluation. Consistent, precise definition and use of terms is critical for biomarker evaluation because it is a topic important across many disciplines and has been for several decades. The committee has attempted to be consistent with the spirit of previous efforts at standardizing the language used with reference to biomarker evaluation, and clarifies several definitions where there is overlap or potential for confusion. Several of the definitions used in the report summary (see Box 1-1 below) deserve further discussion. A definition of risk biomarker, used in the statement of task, is also defined in Box 1-1.

The definition of the term "biomarker" itself is not controversial. The definition provided by the Biomarkers Definitions Working Group is widely used, and other definitions do not differ fundamentally. The *Cancer Biomarkers* report presented two tables showing uses of biomarkers in clinical and drug development settings (see Tables 1-1 and 1-2). The committee viewed results from imaging tests as biomarkers because they are measurements that indicate normal biological processes, predict risk for disease, and monitor pathogenic processes and pharmacologic responses to therapeutic interventions. The committee also viewed genes, genetic signatures, and genetic mutations as biomarkers. While these are typically not modifiable, they do fulfill the Biomarkers Definitions Working Group definition of a biomarker, as they indicate normal biological processes, pathogenic processes, or pharmacologic responses.

The statement of task for this study cites "risk biomarkers" for chronic disease. The committee defines a risk biomarker as a biomarker that

[4] In legislation, the term National Academy of Sciences refers to the whole of the National Academies.

BOX 1-1
Important Definitions

Analytical Validation: "assessing [an] assay and its measurement performance characteristics, determining the range of conditions under which the assay will give reproducible and accurate data."[a]

Biomarker: "a characteristic that is objectively[b] measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a[n] . . . intervention."[c] Example: cholesterol level.

Chronic Disease: a culmination of a series of pathogenic processes in response to internal or external stimuli over time that results in a clinical diagnosis/ailment and health outcomes. Example: diabetes.

Clinical Endpoint: "a characteristic or variable that reflects how a patient [or consumer] feels, functions, or survives."[c] Example: death.

Fit-for-Purpose: being guided by the principle that an evaluation process is tailored to the degree of certainty required for the use proposed.

Qualification: "evidentiary process of linking a biomarker with biological processes and clinical endpoints."[d]

Surrogate Endpoint: "a biomarker that is intended to substitute for a clinical endpoint. A surrogate endpoint is expected to predict clinical benefit (or harm or lack of benefit or harm) based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence."[c] Example: blood pressure for trials of several classes of antihypertensive drugs.[e]

NOTES: [b] The committee defines "objectively" to mean "reliably and accurately." [e] Please see Chapter 2 for discussion of this biomarker.
SOURCES: [a] Wagner (2002); [c] Biomarkers Definitions Working Group (2001); and [d] Wagner (2008).

indicates a risk factor for a disease. In other words, it is a biomarker that indicates a component of an individual's level of risk for developing a disease or level of risk for developing complications of a disease. The committee viewed risk biomarkers as a subset of risk factors. Risk factors are variables that correlate with incidence of a disease or condition. Risk factors include social and environmental factors in addition to biological factors. Risk biomarkers are also to be distinguished from biomarkers of exposure used in toxicology, which were defined by the National Research Council as "the chemical or its metabolite or the product of an interaction between a chemical and some target molecule or cell that is measured in a compartment in an organism" (NRC, 2006; WHO, 2001). In its *Guidance for Industry: Evidence-Based Review System for the Scientific Evaluation of Health Claims*, CFSAN defined risk biomarkers as "biologi-

TABLE 1-1 Use of Biomarkers in Chronic Disease Patient Care

Clinical Biomarker Use	Clinical Objective
Disease risk stratification	Assess the likelihood that the disease will develop (or recur)
Prevention	Identify and track risk factors
Screening[a]	Detect and treat early-stage disease in the asymptomatic population
Diagnosis	Definitively establish the presence of disease
Classification[b]	Classify patients by disease subset
Prognosis	Predict the probable outcome of disease to determine the aggressiveness of treatment
Prediction/treatment stratification[b]	Predict response to particular therapies and choose the drug that is mostly likely to yield a favorable response in a given patient
Therapy-related risk management[a]	Identify patients with a high probability of adverse effects of a treatment
Therapy monitoring[c]	Determine whether a therapy is having the intended effect on a disease and whether adverse effects arise
Surveillance	Early detection and treatment of advancing disease or complications

NOTES: [a] In toxicology, biomarkers of exposure help predict an individual's risk of suffering consequences from exposure to a foreign substance. Exposure biomarkers are a subset of these two categories. [b] Companion diagnostic biomarkers include features from several of these categories. These tests identify whether an individual's molecular profile associated with a disease pathophysiology is likely to respond favorably to a particular therapeutic. Examples include KRAS–cetuximab, HER-2–herceptin, and estrogen receptor status–tamoxifen. [c] Dose optimization is a subset of this category.
SOURCE: Adapted from IOM (2007).

TABLE 1-2 Use of Biomarkers in Drug Development

Biomarker Use	Drug Development Objective
Target validation	Demonstrate that a potential drug target plays a key role in the disease process
Early compound screening	Identify compounds with the most promise for efficacy and safety
Pharmacodynamic assays	Determine drug activity; select dose and schedule
Patient selection	In clinical trials, patient selection (inclusion/exclusion) by disease subset or probability of response/adverse events
Surrogate endpoint	Use of a short-term outcome measure in place of the long-term primary endpoint to determine more quickly whether the treatment is efficacious and safe in drug regulatory approval

SOURCE: IOM (2007).

cal indicators that signal a changed physiological state that is associated with the risk of a disease" (CFSAN, 2009). This definition is narrower than the committee's because it would seem not to include genetic risk factors and other situations that may be present in an individual from birth. Many risk biomarkers are not modifiable in beneficial ways, even when only ones indicating changed physiological states are considered. It is important to note that while some so-called risk biomarkers have been used as surrogate endpoints, risk biomarkers are not surrogate endpoints unless they are determined to be supported for use as such for a defined context of use through use of the biomarker evaluation framework and expert panel as described in Recommendations 1 and 2.

The definition of "surrogate endpoint" is critical for clear communication and transparency in regulatory processes. Several definitions of surrogate endpoint have been used. Table 1-3 shows definitions that have appeared in regulations and other regulatory documents. Table 1-4 shows literature definitions.

TABLE 1-3 Regulatory Definitions of Surrogate Endpoint

Source	Definition
57 *FR* 13234–13242 (1992)[a]	A surrogate end point, or "marker," is a laboratory measurement or physical sign that is used in therapeutic trials as a substitute for a clinically meaningful endpoint that is a direct measure of how a patient feels, functions or survives and is expected to predict the effect of the therapy.
FDAMA (Food and Drug Administration Modernization Act) 1997 USC Section 504(b)(1)	…a surrogate endpoint that is reasonably likely to predict clinical benefit.
Title 21 – Food and Drugs 21 C.F.R. 314 Section 314.510[b]	…a surrogate endpoint that is reasonably likely, based on epidemiologic, therapeutic, pathophysiologic, or other evidence, to predict clinical benefit.
Guidance for Industry: Evidence-based review system for the scientific evaluation of health claims[c]	Surrogate endpoints are risk biomarkers that have been shown to be valid predictors of disease risk and therefore may be used in place of clinical measurements of the onset of the disease in a clinical trial.

SOURCES: [a] New drug, antibiotic and biological drug product regulations: accelerated approval. Proposed Rule. 57 *Federal Register* 13234–13242 (1992). [b] Food and Drug Modernization Act of 1997, 21 USC section 506(b)(1) (1997). Title 21—Food and Drugs, 21 CFR 314 Section 314.510 (2008) [http://frwebgate5.access.gpo.gov/cgi-bin/TEXTgate.cgi?WAISdocID=026369143256+87+1+0&WAISaction=retrieve]. [c] http://www.cfsan.fda.gov/~dms/hclmgui6.html.

TABLE 1-4 Literature Definitions of Surrogate Endpoint

Source	Definition
Biomarkers Definitions Working Group (2001)	A biomarker that is intended to substitute for a clinical endpoint. A surrogate endpoint is expected to predict clinical benefit (or harm or lack of benefit or harm) based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence.
Guide to Clinical Trials (Spilker, 1991)	The ideal surrogate endpoint is a disease marker that reflects what is happening with the underlying disease. The relationship between the marker and the true endpoint is important to establish. After this is done, the validity of data based on how the marker is affected by a medicine or other treatment can be translated into a valid statement about the disease and true endpoint.
Prentice (1989)[a]	A response variable for which a test of the null hypothesis of no relationship to the treatment groups under comparison is also a valid test of the corresponding null hypothesis based on the true endpoint.
Temple (1995)[a]	A surrogate endpoint of a clinical trial is a laboratory measurement or a physical sign used as a substitute for a clinically meaningful endpoint that measures directly how a patient feels, functions or survives. Changes induced by a therapy on a surrogate endpoint are expected to reflect changes in a clinically meaningful endpoint.
Johnston (1999)[a]	A surrogate outcome measure is simply one that is used in place of a clinical endpoint ... an adequate surrogate measure must not only correlate with the clinical endpoint, but it must be predictive of the clinical endpoint in the presence of the intervention under study.
Baker et al. (2005)[a]	A surrogate endpoint is defined as a measure or indicator of a biological process that is obtained sooner, at less cost or less invasively than a true endpoint of health outcome, and is used to make conclusions about the effect of an intervention on the true endpoint.
Grimes and Schulz (2005)[a]	A valid surrogate endpoint must both correlate with and accurately predict the outcome of interest.
Gluud et al. (2007)[a]	A surrogate outcome measure is a laboratory measurement, a physical sign, or any other intermediate substitute that is able to predict a treatment response on a clinically meaningful outcome measure.
Pryseley et al. (2007)[a]	A surrogate for a true endpoint is an endpoint that can be used in lieu of the true endpoint to assess treatment benefits. That is, the effect of the treatment on the surrogate endpoint should reliably predict the effect of the treatment on the true endpoint.
Gobburu (2009); Lathia et al. (2009)	A biomarker that is intended to substitute for a clinical endpoint.

NOTE: [a] See also the Shi and Sargent (2009) compilation of these definitions.

There are a few common features in the overwhelming majority of the surrogate endpoint definitions. First, a surrogate endpoint is meant to substitute for a clinically meaningful endpoint. Second, the surrogate endpoint needs to predict change in those clinical outcomes given an intervention. The last definition in Table 1-4 appears to be for a proposed surrogate endpoint, not for one that has already been determined to satisfy the requirements of a true surrogate endpoint. The committee views this definition as being too inclusive to be accurate. This definition is not consistent with consensus and most regulatory definitions of surrogate endpoint. The last definition in Table 1-3 is the definition used by CFSAN for review of health claims that industry submits for inclusion in food labeling. The citation given in the guidance document is for Spilker's *Guide to Clinical Trials* (shown in Table 1-4; 1991); however, the definition in the guidance document is not consistent with the one it cites. In Dr. Spilker's more recent book, *Guide to Drug Development: A Comprehensive Review and Assessment* (2009), the Biomarkers Definitions Working Group definition is used. The CFSAN definition does not include a critical component of the definition of surrogate endpoints: the ability to predict clinical benefit or harm of an intervention based on a change in the surrogate endpoint. The use of the word "valid" in this definition is also ambiguous, as will be discussed below. Finally, the CFSAN definition accounts only for use of surrogate endpoints in clinical trials and does not allow for use in observational studies. The Biomarkers Definitions Working Group's definition takes into account uses of surrogate endpoints in observational studies.

There are a number of other important concepts to understand when considering surrogate endpoints. The Prentice criteria are succinctly summarized in two parts: correlation and capture. Under correlation, the surrogate endpoint must be statistically correlated to the clinical endpoint. In other words, the surrogate endpoint should have prognostic value relative to the clinical endpoint. Under capture, an intervention's entire effect on the clinical endpoint should be explained by the intervention's effect on the surrogate endpoint. In other words, the surrogate endpoint should account for all of an intervention's effects; the surrogate endpoints should be a perfect proxy for the effect of an intervention on the recipient's risk of important clinical outcomes (Desai et al., 2006; Prentice, 1989).

The terms "clinical endpoint" and "true endpoint" are sometimes used interchangeably. The definition of clinical endpoint given in Box 1-1 is widely accepted and consistently used, while the term true endpoint is broader and ill defined. To some, only all-cause mortality is a true endpoint. In practice, however, a trial's true endpoint is defined by the experimenters. It can be mortality due to the disease being studied, failure of the treatment (which can be defined in several ways), time to progres-

sion, or something else. Sometimes, a surrogate endpoint in one study can be the clinical endpoint in another study. Practically, the true endpoint is the endpoint for which a surrogate endpoint is sought. Myocardial infarction (MI) is an example. Because an MI in a person outside the hospital is detected from symptoms, it is a plausible clinical endpoint. However, it should be acknowledged that a significant element of the importance of MI derives from both the fact that it is a biomarker for risk of future events (death, heart failure) and that it requires objective biomarker measurements for the diagnosis.

The term "validation" encompasses many different aspects of biomarker development. In the statistics literature, validation means what other fields term "qualification." Validation and analytical validation are often used interchangeably, as are clinical validation and qualification. Clinical utility is often used interchangeably with utilization. In this report, the committee uses validation and analytical validation interchangeably, qualification but not clinical validation, and utilization but not clinical utility.

Correct definition of the terms food, substance, disease, and drug are important for understanding FDA regulations. Food is defined as (1) articles used for food or drink for humans or other animals, (2) chewing gum, and (3) articles used for components of any such article.[5] As was noted in the summary of this report, however, the committee has been more explicit in its definition: the term "food" is inclusive of foods consumed as part of meals and snacks, dietary supplements, and components contained in them (nutrients, other bioactive substances). A substance is "a specific food (tomato) or component of food (lycopene), whether in conventional food or dietary supplement form"[6] (Trumbo and Ellwood, 2009). A disease or health-related condition is "Damage to an organ, part, structure, or system of the body such that it does not function properly (e.g., CHD), or a state of health leading to such dysfunctioning (e.g., hypertension)"[7] (Trumbo and Ellwood, 2009). A drug is defined as "articles intended for use in the diagnosis, cure, mitigation, treatment, or prevention of disease" and "articles (other than food) intended to affect the structure or any function of the body of man or other animals"[8] (FDA, 2002). The term "intervention" refers to any drug, device, biologic, behavioral modification, nutritional modification, lifestyle modification, or other treatment intended to improve health.

[5] FDCA, Sec. 201(II)(f).
[6] 21 C.F.R. 101.14(a)(2).
[7] 21 C.F.R. 101.14(a)(5).
[8] FDCA, Sec. 201(g)(1).

RELATED IOM WORK

The committee views this report as building on and supporting the recommendations of several previous committees. In particular, the committee would like to reemphasize the recommendations of the report on *Cancer Biomarkers* (Box B-1) and the report on *The Future of Drug Safety* (Box B-2). The recommendations from both of these reports are included in Appendix B. *Cancer Biomarkers* grouped its recommendations into three categories: (1) methods, tools, and resources needed to discover and develop tools for cancer; (2) guidelines, standards, oversight, and incentives needed for biomarker development; and (3) methods and processes needed for clinical evaluation and adoption. Government agencies, academics, healthcare practitioners, industrial stakeholders, and the Institute of Medicine have been working to explore and implement changes that reflect the needs identified in the recommendations. As mentioned earlier, the current report was requested by the FDA as a path forward on recommendation 6 from the *Cancer Biomarkers* report.

The recommendations from *The Future of Drug Safety* were grouped into categories: organizational culture, science and expertise, regulation, communication, and resources. Following the release of the report in 2007, the Food and Drug Administration Amendments Act was passed. It reauthorized a number of key pieces of legislation important for increasing drug safety and expanded FDA responsibilities and capabilities to respond to a number of *The Future of Drug Safety* report's recommendations (FDA, 2009b). In 2009, the FDA published a table describing the significant progress made on implementation of the IOM recommendations (FDA, 2009c).

FRAMEWORK OF THE REPORT

The framework of the report follows the statement of task and the committee's recommendations. Chapter 2 reviews previous biomarker and surrogate endpoint evaluation processes. Chapter 3 presents the committee's recommended biomarker evaluation framework. Chapter 4 contains the case studies that exemplify use of the biomarker evaluation process. Finally, Chapter 5 describes data collection and data infrastructure needs to support the FDA's work.

REFERENCES

Adams, C. P., and V. V. Brantner. 2006. Estimating the cost of new drug development: Is it really $802 million? *Health Affairs* 25(2):420–428.

Baker, S. G., G. Izmirlian, and V. Kipnis. 2005. Resolving paradoxes involving surrogate end points. *Journal of the Royal Statistical Society Series A* 168(4):753–762.

Biomarkers Definitions Working Group. 2001. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology and Therapeutics* 69(3):89–95.

CFSAN (Center for Food Safety and Applied Nutrition). 2009. *Guidance for industry: Evidence-based review system for the scientific evaluation of health claims*. http://www.cfsan.fda.gov/~dms/hclmgui6.html (accessed February 25, 2009).

Desai, M., N. Stockbridge, and R. Temple. 2006. Blood pressure as an example of a biomarker that functions as a surrogate. *AAPS Journal* 8(1):E146.

FDA (Food and Drug Administration). 2002. *Is it a cosmetic, a drug, or both? (Or is it soap?)* http://www.fda.gov/Cosmetics/GuidanceComplianceRegulatoryInformation/ucm074201.htm (accessed October 2, 2009).

FDA. 2004a. *Challenge and opportunity on the critical path to new medical products*. http://www.fda.gov/oc/initiatives/criticalpath/whitepaper.html (accessed December 10, 2008).

FDA. 2004b. *Challenge and opportunity on the critical path to new medical products.* Washington, DC: FDA.

FDA. 2006. *Critical Path opportunities list*. http://www.fda.gov/oc/initiatives/criticalpath/reports/opp_list.pdf (accessed December 10, 2008).

FDA. 2008. *Key FDA Critical Path activities under way in 2007*. http://www.fda.gov/oc/initiatives/criticalpath/report2007.html (accessed December 10, 2008).

FDA. 2009a. *CFSAN – What we do*. http://www.fda.gov/AboutFDA/CentersOffices/CFSAN/WhatWeDo/default.htm (accessed October 1, 2009).

FDA. 2009b. *Regulatory information: Food and Drug Administration Amendments Act (FDAAA) of 2007*. http://www.fda.gov/RegulatoryInformation/Legislation/FederalFoodDrugandCosmeticActFDCAct/SignificantAmendmentstotheFDCAct/FoodandDrugAdministrationAmendmentsActof2007/default.htm (accessed November 20, 2009).

FDA. 2009c. *Report to Congress: Changing the future of drug safety: FDA initiatives to strengthen and transform the drug safety system*. http://www.fda.gov/downloads/Safety/SafetyofSpecificProducts/UCM184046.pdf (accessed November 20, 2009).

Gluud, C., J. Brok, Y. Gong, and R. L. Koretz. 2007. Hepatology may have problems with putative surrogate outcome measures. *Journal of Hepatology* 46(4):734–742.

Gobburu, J. V. S. 2009. Biomarkers in clinical drug development. *Clinical Pharmacology and Therapeutics* 86(1):26–27.

Grimes, D. A., and K. F. Schulz. 2005. Surrogate end points in clinical research: Hazardous to your health. *Obstetrics and Gynecology* 105(5 Pt 1):1114–1118.

IOM (Institute of Medicine). 2006. *Developing biomarker-based tools for cancer screening, diagnosis, and treatment: The state of the science, evaluation, implementation, and economics.* Washington, DC: The National Academies Press.

IOM. 2007. *Cancer biomarkers: The promises and challenges of improving detection and treatment*. Washington, DC: The National Academies Press.

Johnston, K. 1999. What are surrogate outcome measures and why do they fail in clinical research? *Neuroepidemiology* 18(4):167–173.

Lathia, C. D., D. Amakye, W. Dai, C. Girman, S. Madani, J. Mayne, P. MacCarthy, P. Pertel, L. Seman, A. Stoch, P. Tarantino, C. Webster, S. Williams, and J. A. Wagner. 2009. The value, qualification, and regulatory use of surrogate end points in drug development. *Clinical Pharmacology and Therapeutics* 86(1):32–43.

NRC (National Research Council). 2006. *Human biomonitoring for environmental chemicals*. Washington, DC: The National Academies Press.

Prentice, R. L. 1989. Surrogate endpoints in clinical trials: Definition and operational criteria. *Statistics in Medicine* 8(4):431–440.

Pryseley, A., A. Tilahun, A. Alonso, and G. Molenberghs. 2007. Information-theory based surrogate marker evaluation from several randomized clinical trials with continuous true and binary surrogate endpoints. *Clinical Trials* 4(6):587–597.

Sardana, G., B. Dowell, and E.P. Diamandis. 2008. Emerging biomarkers for the diagnosis and prognosis of prostate cancer. *Clinical Chemistry* 54(12):1951–1960.

Shi, Q., and D. J. Sargent. 2009. Meta-analysis for the evaluation of surrogate endpoints in cancer clinical trials. *International Journal of Clinical Oncology* 14(2):102–111.

Spilker, B. 1991. Surrogate endpoints. In *Guide to clinical trials*. New York: Raven Press. Pp. 608–609.

Spilker, B. 2009. Surrogate endpoints and biomarkers. In *Guide to drug development: A comprehensive review and assessment*. Philadelphia, PA: Lippincott Williams and Wilkins.

Temple, R. J. 1995. A regulatory authority's opinion about surrogate endpoints. In *Clinical measurement in drug evaluation*, edited by W. S. Nimmo and G. T. Tucker. New York: John Wiley and Sons. Pp. 3–22.

Trumbo, P., and K. Ellwood. 2009. *Developing a framework for biomarker qualification for chronic disease*. Paper presented at Institute of Medicine Committee on Qualification of Biomarkers as Surrogate Endpoints for Chronic Disease Risk, Washington, DC, January 12.

Wagner, J. A. 2002. Overview of biomarkers and surrogate endpoints in drug development. *Disease Markers* 18(2):41–46.

Wagner, J. A. 2008. Strategic approach to fit-for-purpose biomarkers in drug development. *Annual Review of Pharmacology and Toxicology* 48:631–651.

WHO (World Health Organization). 2001. Biomarkers in risk assessment: Validity and validation. Environmental health criteria 222. http://www.inchem.org/documents/ehc/ehc/ech222.htm (accessed March 16, 2010).

2

Review: Evaluating and Regulating Biomarker Use

INTRODUCTION

The context within which this study is set has developed from the contributions of various scientific fields, industries, and government bodies. From toxicology to cardiology, from the food industry to the drug industry, and from the Food and Drug Administration (FDA) to the federal courts, biomarkers and the scientific evidence needed to substantiate their use have been topics of discussion for several decades. Along with a brief review of biomarker evaluation methods and their uses, this chapter seeks to describe critical areas of background information so that readers from different fields can gain a more comprehensive understanding of the policy and regulatory issues with respect to biomarkers.

Methods for evaluation of biomarkers and surrogate endpoints have been reviewed successfully and systematically in the recent past (Lassere, 2008; Shi and Sargent, 2009). This chapter will direct the readers toward appropriate reviews, and it will discuss the evolution of thinking at the FDA—focusing on the Center for Food Safety and Applied Nutrition (CFSAN), in particular—regarding surrogate endpoints. It will also discuss the evolution in thinking in academic and industry communities, to a lesser extent. The contents of this chapter are as follows:

- Use of biomarkers in areas as diverse as scientific research, medical practice, product development, and public health policy
- Use of biomarkers as surrogate endpoints
- Evaluation frameworks proposed from academia and industry

- The broader context of biomarker and surrogate endpoint evaluation by the FDA, including the legal and regulatory basis for claims made on CFSAN-regulated products

Examples are included on blood pressure as a surrogate endpoint, HIV/AIDS drug development, arrhythmia suppression interventions, exercise tolerance in congestive heart failure, and kidney toxicity biomarkers.

SURVEY OF BIOMARKER USES

Biomarkers have a wide array of uses in a variety of fields. These fields include medicine, oral health, mental health, nutrition, environmental health, toxicology, developmental biology, and basic scientific research. They are used to study the safety and efficacy of interventions, develop understanding of the mechanisms of disease, make good decisions in clinical care, and guide the policies that impact public health. Table 2-1 gives a list of several categories of biomarker use.

For the uses in Table 2-1, any biomarker would need to be evaluated to ensure that data supporting the biomarker's association with the disease or condition of interest and the analytical validation of the test are adequate for the proposed use. In situations, however, where biomarker data will not or is not yet anticipated to be submitted to the FDA for a regulatory purpose or used by professional societies or other groups for clinical practice guidelines or other decision-making processes impacting public health or the practice of medicine, this may be an informal process. Ideally, evaluations are already done by clinicians, product developers, government regulators, professional societies, and scientists; this report's contribution is to propose a systematic process for biomarker evaluation.

Use of Biomarkers and Surrogate Endpoints for Clinical Efficacy Studies and Formation of Clinical Practice Guidelines

Surrogate endpoints were defined in Chapter 1 and can be found in several locations in Table 2-1. First, they have been used in approvals of products or claims for drugs, biologics, devices, foods, and supplements. This will be discussed further in several subsections of this chapter's section on evolution of regulatory perspectives on surrogate endpoints and in Chapter 5. Second, they have been used in the formulation of clinical practice guidelines. As defined by an Institute of Medicine (IOM) committee in 1990, "practice guidelines are systematically developed statements to assist practitioner and patient decisions about appropriate health care

TABLE 2-1 Categories of Biomarker Use

Use	Description
Discovery	Identification of biochemical, image, or other biomarkers associated with a disease, condition, or behavior of interest; biomarkers identified may be screened for many potential uses, including as a target for intervention to prevent, treat, or mitigate a disease or condition
Early product development	Biomarkers used for target validation, compound screening, pharmacodynamic assays, safety assessments, and subject selection for clinical trials, and as endpoints in early clinical screening (i.e., phase I and II trials)
Surrogate endpoints for claim and product approvals	Biomarkers used for phase III clinical testing and biomarkers used to substantiate claims for product marketing
Clinical endpoints	Biomarkers used as endpoints for clinical trials that measure how a patient feels, functions, or survives; for example, measures of depression, blindness, and muscle weakness are biomarkers that may be used as clinical endpoints
Clinical practice	Biomarkers used by clinicians for uses such as risk stratification, disease prevention, screening, diagnosis, prognosis, therapeutic monitoring, and posttreatment surveillance
Clinical practice guidelines	Biomarkers used to make generalized recommendations for healthcare practitioners in the areas of risk stratification, disease prevention, treatment, behavior/lifestyle modifications, and more
Comparative efficacy and safety	Biomarkers used in clinical studies looking at the relative efficacy, safety, and cost effectiveness of any or all interventions used for a particular disease or condition, including changes in behavior, nutrition, or lifestyle; these studies are a component of comparative effectiveness research
Public health practice	Biomarkers used to track public health status and make recommendations for prevention, mitigation, and treatment of diseases and conditions at the population level

for specific clinical circumstances" (IOM, 1990). Clinical practice guidelines and the systematic reviews that inform them are the subjects for two current IOM studies;[1] the reports are expected in 2011. A guideline regarding treatment of a particular disease may identify target levels for specific biomarkers. In order to arrive at a recommendation for a particular biomarker level, clinical trial and observational data must be evaluated. It is possible that more trials will measure a particular surrogate endpoint in addition to or rather than the clinical endpoint of interest. In these cases, it may be desirable to include data from trials that did not measure the clinical endpoints of interest in the systematic reviews.

It is useful to mention that professional societies play an essential role in helping stakeholders understand the best ways to use biomarker-related information in clinical practice. One way in which professional societies assist in the understanding and use of biomarker data is through the promulgation of clinical practice guidelines. The committee recognized that clinical practice guidelines could use the committee's proposed biomarker evaluation framework in reaching decisions. Other methods of rigorous, systematic review, including the Cochrane Collaboration, may also be valuable in assessing the evidence associated with clinical practice guidelines. One consideration that bodies involved in the work of determining the best clinical practice guideline may need to make is that of cost effectiveness. The committee viewed this topic as being beyond the statement of task for this study and well studied elsewhere, but the committee recognizes that comparisons of interventions looking at the number of quality-adjusted life-years gained through use of an intervention or relative to no intervention are useful.

The IOM recently released a report, *Initial National Priorities for Comparative Effectiveness Research* (IOM, 2009c), which identified six characteristics of comparative effectiveness research, or CER (Box 2-1). In general, use of surrogate endpoints in CER would not fulfill the fourth characteristic of comparative effectiveness research, as identified in the report (IOM, 2009c). Quoted below is the report's description of this characteristic of CER:

> **CER measures outcomes—both benefits and harms—that are important to patients.**
>
> The committee is using the term "effectiveness" in reference to the extent to which a specific intervention, procedure, regimen, or service does what it is intended to do when used under *real-world* circumstances.

[1] Standards for Developing Trustworthy Clinical Practice Guidelines (http://www8.nationalacademies.org/cp/projectview.aspx?key=49125) and Standards for Systematic Reviews of Clinical Effectiveness Research (http://www8.nationalacademies.org/cp/projectview.aspx?key=49124).

BOX 2-1
Characteristics of Comparative Effectiveness Research (CER)

1. CER has the objective of directly informing a specific clinical decision from the patient perspective or a health policy decision from the population perspective.
2. CER compares at least two alternative interventions, each with the potential to be "best practice."
3. CER describes results at the population and subgroup levels.
4. CER measures outcomes—both benefits and harms—that are important to patients.
5. CER employs methods and data sources appropriate for the decision of interest.
6. CER is conducted in settings that are similar to those in which the intervention will be used in practice.
7. CER has the objective of directly informing a specific clinical decision from the patient perspective or a health policy decision from the population perspective.
8. CER compares at least two alternative interventions, each with the potential to be "best practice."
9. CER describes results at the population and subgroup levels.
10. CER measures outcomes—both benefits and harms—that are important to patients.
11. CER employs methods and data sources appropriate for the decision of interest.
12. CER is conducted in settings that are similar to those in which the intervention will be used in practice.

SOURCE: IOM (2009c).

This can be contrasted with "efficacy," which is the extent to which an intervention produces a beneficial result under controlled conditions (Cochrane, 1971; Higgins and Green, 2008). This implies an important distinction between much clinical research and CER, in that CER places high value on external validity, or the ability to generalize results to real-world decision making. Harms or risks of unintended consequences are also outcomes of interest, because they influence the net benefits of an intervention. Including and giving weight to patient-reported outcomes is particularly important for CER studies in which patient ratings of effectiveness or adverse events may differ from clinical measures. Finally, resource utilization may be highly relevant to net benefits when comparing the full clinical course of interventions over time. Cost-effectiveness analysis is a useful tool of CER, allowing evaluation of the full range of

> treatment outcomes in relationship to the difference in costs. Robust evidence of comparative clinical effectiveness is a building block necessary for resource allocation decisions. Moreover, just as clinical effects may vary in different settings, costs vary as well, so a given set of cost-effectiveness results is often not generalizable. (IOM, 2009c)

Comparative effectiveness research is meant to fill gaps in evidence that prevent comparison of available treatments (IOM, 2009c) with a focus on outcome measurements that are tangible to the person rather than biomarkers or putative surrogate endpoints. Occasionally, it may be impractical for many of these studies to examine clinical endpoints; careful selection of surrogate endpoints after significant interaction with patient groups and expert investigators would be necessary. Finally, surrogate endpoints can be found in public health practice when there is a need to estimate the health of populations or short-term impacts of longer-term programs for prevention, treatment, or mitigation of infectious or chronic diseases when health outcomes important to patients cannot be measured. For example, reporting to stakeholders about interventions to decrease diseases and conditions of importance in the population, such as stroke or heart attack, may be done by measuring and reporting blood pressure as a surrogate for the desired improvement in health status, although measuring health outcomes important to patients such as stroke or quality of life would be preferable as guidance to public health interventions unless such measures were deemed impractical.

Surrogate Endpoints: Successes

The most widely discussed use of surrogate endpoints is in phase III clinical studies used to support applications for new drugs, biologics, and devices and to support claims on foods and supplements. In his presentation to the committee during its April public workshop, Dr. Robert Temple of the Center for Drug Evaluation and Research (CDER) at the FDA outlined the reasons why researchers and clinicians use surrogate endpoints (Temple, 2009).

These reasons include when the clinical endpoint is rare or takes years to develop; when the surrogate endpoints seem to be obviously linked to the clinical endpoint of interest (e.g., tumor size in cancer or maintenance of regular heart rhythm in arrhythmia patients); and when other treatments exist, to alleviate the difficulties of conducting trials when a new intervention must be proven as non-inferior to existing treatments. In addition, although it may be possible to use a clinical endpoint in a population at high risk for the disease or condition, studying a population at relatively lower risk using the clinical endpoint may be too burdensome

since the number of subjects required would be very large. Dr. Temple noted that the idea of a surrogate endpoint is to enable faster, smaller, more efficient clinical trials that can address urgent needs and facilitate the advancement of medicine.

Two notable successes of the use of surrogate endpoints are discussed in the next sections: blood pressure and HIV-1 RNA. The first example details the history of the evaluation of blood pressure as a surrogate endpoint. It may be surprising to readers that blood pressure as a surrogate endpoint for cardiovascular disease endpoints was hotly debated for decades before reaching its current status. Still, there is no broad agreement that blood pressure is a universal surrogate endpoint (Carter, 2002; Psaty et al., 1996). Even though these examples describe successful use of surrogate endpoints, important caveats are also described. Dr. Temple and others have noted surprises and mistakes in the selection and use of surrogate endpoints, and so several examples of these are discussed after the sections on blood pressure and HIV-1 RNA.

Blood Pressure

Blood pressure is often looked to as an exemplar surrogate endpoint for cardiovascular mortality and morbidity due to the levels and types of evidence that support its use. More than 75 antihypertensive agents in more than 9 therapeutic classes demonstrate the wide availability of agents to treat hypertension (Israili et al., 2007). Although new antihypertensive drugs are approved on the basis of blood pressure reductions, blood pressure's history as a surrogate endpoint is unusual in that many drugs used to treat hypertension (thiazides, methyldopa, reserpine, hydralazine, guanethidine) were approved prior to the FDA's effectiveness requirement or the availability of clinical trial data supporting the impact of blood pressure control on cardiovascular outcomes (Desai et al., 2006).

The status of blood pressure as a surrogate endpoint for cardiovascular disease endpoints was debated for decades (Perry et al., 1978). Even as one of the most well-established surrogate endpoints, an effect on blood pressure may not fully capture the benefit—or risk—of an intervention.

Although some issues are still outstanding, the benefits of blood pressure control are mostly well understood due to comprehensive epidemiologic and clinical trial evidence. Hypertension has been identified as the most common risk biomarker for cardiovascular morbidity and mortality, with a World Health Organization report suggesting that hypertension is the single most important preventable cause of premature death in developed countries (Ezzati et al., 2002). Data suggest that in the United States, hypertension is responsible for 35 percent of myocardial infarctions

and strokes, 49 percent of episodes of heart failure, and 24 percent of premature deaths (Wolff and Miller, 2007). Hypertension affects one in four U.S. adults, but the majority of those affected remain either untreated or undertreated in spite of the substantial health benefits gained from modest blood pressure reductions (Wang and Vasan, 2005).

Epidemiological, clinical trial data Williams (2005) suggested that the blood pressure–cardiovascular outcomes relationship is substantiated by one of the strongest evidence bases in clinical medicine. Epidemiologic studies consistently demonstrate the relationship between blood pressure and cardiovascular mortality and morbidity, including one meta-analysis of nine studies that demonstrated an association between diastolic blood pressure and coronary heart disease and stroke in 420,000 subjects (MacMahon et al., 1990). Observational studies have also demonstrated the robustness of blood pressure's relationship to heart disease in adults; despite different assessment parameters (systolic alone, diastolic alone, or systolic and diastolic), the relationship is maintained (Desai et al., 2006). This relationship has also been confirmed in diverse populations, including different genders, adult age groups, and race/ethnicities. In children, this relationship does not hold (Brady and Feld, 2009).

Both placebo- and active-controlled clinical trials conducted in the past three to four decades have demonstrated that pharmacologic reductions in blood pressure reduce cardiovascular mortality and morbidity (Desai et al., 2006). While earlier trials compared hypertension agents against placebo, the growing evidence base supporting the benefit of hypertension therapy necessitated head-to-head trials comparing two or more agents, which reduced power of the studies and required much larger numbers of patients to see an effect (Williams, 2005). Many different therapeutic agents—including diuretics, beta blockers, angiotension converting enzyme (ACE) inhibitors, calcium channel blockers, and angiotensin receptor blockers—are approved to lower blood pressure.

Effects of blood pressure-lowering drugs Impact on blood pressure may or may not capture an intervention's entire risk–benefit balance. Different classes of agents, or even agents within a specific class, may have multiple effects, one of which is lowering blood pressure (NHLBI Working Group, 2005). For example, ACE inhibitors are known to have at least 10 pharmacologic effects (Borer, 2004). This notion has generated trials testing whether agents have beneficial effects that go beyond blood pressure lowering. ALLHAT (Antihypertensive and Lipid Lowering Treatment to Prevent Heart Attack Trial) compared the efficacy of four different drug classes (a calcium channel blocker, an ACE inhibitor, an alpha adrenergic blocker, and a diuretic) for initial therapy of hypertension. Study results

demonstrated that three classes of drugs (calcium channel blocker, ACE inhibitor, and diuretic) could not be distinguished for the primary endpoint, coronary heart disease (CHD) mortality and non-fatal myocardial infarction, but the lower cost diuretics were superior in regard to secondary outcomes and should be the preferred first step therapy (ALLHAT Officers and Coordinators, 2002). The alpha adrenergic blocker arm of the trial was dropped because of the significantly higher incidence of combined cardiovascular events in the alpha adrenergic blocker arm compared to the diuretic, including a two-fold relative risk of congestive heart failure compared to the diuretic (ALLHAT Officers and Coordinators, 2000).

Other conclusions have also been drawn from these large, prospective head-to-head comparison trials; some investigators suggest that it is the blood pressure reduction, rather than the specific drug used, that confers cardiovascular benefit (Williams, 2005). In an analysis of 147 randomized trials, investigators found that all classes of blood pressure-lowering drugs have similar effects in reducing coronary heart disease events and strokes for a given level of blood pressure reduction, with the exception of an extra protective effect of beta blockers administered shortly after myocardial infarction and minor protective effect of calcium channel blockers in stroke (Law and Morris, 2009). Although there is still some ambiguity about the use of differing blood pressure agents, the fact that pharmacologically distinct agents have directionally similar effects on cardiovascular outcomes has provided more support for the use of blood pressure as a surrogate endpoint for coronary heart disease and stroke.

Regulatory use of blood pressure as a surrogate endpoint The consistent demonstration that diverse blood pressure-lowering agents confer cardiovascular benefits, as well as the substantial epidemiological data linking hypertension to cardiovascular events, provides the basis for the FDA's use of blood pressure as a surrogate endpoint (Desai et al., 2006; Temple, 1999). However, clear guidance on the use of surrogate endpoints within the FDA is lacking because the Food, Drug, and Cosmetic Act does not specifically state which endpoints—or criteria—can be used for drug approval. Through case law, the FDA has the authority to deny approval of a drug on the basis of its effect on the surrogate endpoint if the surrogate endpoint's clinical value is unknown.[2] In 1992, FDA regulation provided a new method for drug approval on the basis of effects on a surrogate endpoint, called accelerated approval, for serious or life-threatening conditions without available therapy. The regulation stated that drugs could be approved on the basis of surrogate endpoint data if it "is reasonably

[2] *Warner-Lambert* v. *Heckler*, 787 F.2d 147 (3rd Cir. 1986).

likely, based on epidemiologic, therapeutic, pathophysiologic, or other evidence, to predict clinical benefit"[3] and required confirmatory clinical evidence. The regulation also referenced "well-established" surrogates on which drug approval had been based, but did not define well-established endpoints. Temple (1999) noted that "well-established" surrogates would need to be more than "reasonably likely" to predict benefit.

Despite the lack of clarity in the regulations concerning surrogate endpoints, the FDA accepts surrogate endpoints for drug approval and as the basis for authorized health claims. However, different divisions and centers within the FDA accept different surrogate endpoints. For example, the Cardio-Renal Division within the CDER accepts blood pressure reduction as a surrogate endpoint for cardiovascular event reduction, but requires direct clinical benefit measurement for other endpoints, while the Metabolic-Endocrine Division also accepts LDL-C lowering as a surrogate endpoint for cardiovascular events (Borer, 2004). The Metabolic-Endocrine Division also accepts use of glycosylated hemoglobin level and blood glucose control as surrogate endpoints for diabetes control (Borer, 2004). Even so, the FDA has recognized the inadequacy of small six-month trials that address effects of type 2 diabetes mellitus treatments on HbA1c, and now the FDA requires large-scale randomized cardiovascular safety clinical endpoint trials be conducted pre- and post-approval.

Within CFSAN, blood pressure is recognized as a surrogate endpoint for hypertension (FDA, 1999). Hypertension is considered a disease-related health condition. As discussed earlier, hypertension—high blood pressure—is recognized as a strong risk factor for cardiovascular disease. CFSAN has authorized a health claim for low-sodium foods based on the surrogate endpoint–disease-related condition relationship, stating either "diets low in sodium may reduce the risk of high blood pressure, a disease associated with many factors" or "development of hypertension or high blood pressure depends on many factors. [This product] can be part of a low sodium, low salt diet that might reduce the risk of hypertension or high blood pressure."[4]

HIV Drug Development

One of the motivations for the earliest efforts at surrogate endpoint evaluation arose from the acute need for effective therapeutics early in the HIV/AIDS epidemic. The early trials of anti-HIV therapies used progression to AIDS or death as the clinical outcome measures. These studies could be short in some settings, like those in which the effects of the

[3] 21 C.F.R. § 601 (2008).
[4] 21 C.F.R. § 101.74 (2009).

intervention were large and participants had advanced disease (Fischl et al., 1987; Hammer et al., 1997). Studies could also be short when they were large enough so that only a small percentage of patients who progress to advanced disease drove the principal finding (Volberding et al., 1994). However, the latter type of study could produce misleading results in that a small number of patients destined to progress quickly might benefit from an intervention, like AZT monotherapy, while an even larger number might experience no benefit and even positive harm following the conclusion of the study, because of factors like the development of resistance to the drug under study and others with similar mechanisms of action. Such concerns underscored the need for a more rapid means of evaluating the benefit of antiviral therapy that might reflect risk or benefit to a larger proportion of the study population more rapidly.

Early in the AIDS epidemic, it was observed that clinical disease progression was associated with a decline of $CD4^+$ T-lymphocytes (CD4 cells); in the 1990s, a virologic measure that both responded to therapy and predicted outcomes was developed (HIV-1 RNA). The earliest approval of a drug based on a biomarker—didanosine was approved in 1991—used CD4 cell count; however, the development of measurement of plasma HIV-1 RNA by polymerase chain reaction (PCR), which made a direct measurement of viral replication possible, rapidly became the standard endpoint in HIV clinical trials. In the mid-1990s, representatives from industry, drug regulatory agencies, and academia sought to formally evaluate CD4 cell count and HIV-1 RNA as surrogate endpoints for disease progression in clinical trials and in patient management (Hughes et al., 1998).

To evaluate HIV-1 RNA and CD4 cell count as surrogate endpoints, the HIV Surrogate Marker Collaborative Group, a group involving statisticians and clinicians from pharmaceutical companies and government-funded cooperative clinical trials groups, was formed. The HIV Surrogate Marker Collaborative Group undertook a meta-analysis of clinical trials to evaluate treatment-mediated changes in HIV-1 RNA and CD4 cell count as surrogate endpoints (HIV Surrogate Marker Collaborative Group, 2000). The meta-analysis found that HIV-1 RNA and CD4 cell count have independent value as prognostic biomarkers. However, the meta-analysis also found that short-term changes in the values of these biomarkers were not adequate surrogate endpoints for determining the impact of an intervention on long-term clinical endpoints such as progression to AIDS and death (HIV Surrogate Marker Collaborative Group, 2000). Their analysis also showed that changes in HIV-1 RNA explained only about half of the benefit of treatment. However, these results mostly reflected the experience of patients on drug regimens that were not capable of suppressing most patients' viral loads below levels of assay detection.

In 2002, the FDA issued a guidance for industry that advocated the use of HIV-1 RNA in plasma as the primary basis for assessing efficacy of antiretroviral drugs for accelerated and traditional approval, although it had begun approving drugs based on evidence of lower levels of plasma HIV-1 RNA a few years earlier (Behrman, 1999). Additionally, it recommended that "changes in CD4 cell counts be consistent with observed HIV-1 RNA changes when considering approval of an antiretroviral drug" (FDA, 2002). In most cases, approval was based on demonstrations that new drugs, used in combination with existing drugs, were able to suppress virus among patients who had not been previously exposed to therapy and had virus that was sensitive to at least one other agent in the regimen. An important distinction must be made between using HIV-1 RNA as a surrogate for a clinical endpoint in a setting where virus can be fully suppressed and a setting where virus is only partly, and often therefore temporarily, suppressed. Complete viral suppression often leads to durable suppression, perhaps because of the lower risk of development of viral resistance mutations in patients without replicating virus. Tolerable drugs that produce durable suppression are likely to benefit patients because such suppression is associated with steady improvements in CD4 and reduced risk of clinical events associated with HIV infection.

The value of HIV-1 RNA as a surrogate in settings where suppression of HIV-1 RNA is partial is much more problematic and contingent on context, because partial HIV suppression invites development of new drug resistance mutations that limit the future usefulness of the drugs under study and similar drugs. Therefore a drug that induces a temporary reduction in HIV-1 RNA, while perhaps valuable in reducing risk of clinical disease over a short interval, may reduce the possibility of later construction of a durable three-drug regimen. Such loss of future drug options is an important consequence of drug treatment that is not captured by plasma HIV-1 RNA levels (Jiang et al., 2003). Another important factor is viral fitness, which is affected by treatment and may also be relevant for long-term outcomes (Deeks and Martin, 2007).

As a consequence the use of HIV-1 RNA as a surrogate for clinical endpoint in settings where viral suppression is not complete has not been supported with evidence and probably cannot be. As mentioned above, the relative benefit of different degrees of partial HIV suppression are highly context specific and dependent on the availability of other drugs. De Gruttola et al. (2006), in a discussion of the approval of tipranavir in exactly such a context, recommended that only complete suppression of plasma HIV-1 RNA be used in such studies, and that partial suppression endpoints not be used in clinical trials.

Historically, it is important to note that the FDA's guidance to industry occurred prior to the approval of newer types of antiretroviral drugs

that use different mechanisms than those formally evaluated in the meta-analysis (Hughes, 2005). More potent antiretroviral drugs, which can fully suppress HIV-1 viral load, have since become standard of care. This suggests that although HIV-1 RNA has become the primary endpoint to determine efficacy in many antiretroviral trials, collection of additional and longer term information that relates to both risk and benefit—especially in studies of newer types of antiretroviral drugs—is warranted.

In conclusion, the rapid development of HIV drugs in the 1990s was enabled through the use of surrogate endpoints. While this use of surrogate endpoints inspired the creation of the Critical Path Initiative, the process of biomarker evaluation used was not systematic and so was not easily translated into other disease areas. Nonetheless, the success of this effort to speed approvals of HIV drugs highlighted the value that a systematic biomarker evaluation process could have for drug regulation in general.

Cautionary Statements Regarding the Use of Surrogate Endpoints

Remarkably, the cautionary voices speaking about the risks of using surrogate endpoints have been repeating the same messages for 20 years. What has been changing is the continually increasing amount of data supporting their arguments. In 1989, Ross Prentice initiated the conversation about surrogate endpoints with his influential paper, which provided a statistical definition of a surrogate endpoint. In this paper, he wrote, "I am somewhat pessimistic concerning the potential of the surrogate endpoint concept" (Prentice, 1989). This statement was made in acknowledgment of the hope, already palpable, that a surrogate endpoint, once shown useful for one intervention, would be extensible to other interventions and that relative reductions in one risk factor would be comparable to others for a given clinical endpoint.

Editorials in the early 1990s looked at the rapid advances—and mistakes—enabled through use of surrogate endpoints at the beginning of the HIV/AIDS epidemic (Cotton, 1991; De Gruttola et al., 1997; Holden, 1993; Lagakos and Hoth, 1992). The potential benefits and hazards of the use of surrogate endpoints have been understood since the beginning of this discussion. In 1991, Cotton noted several standing questions in relation to use of surrogate endpoints in the treatment of HIV/AIDS. Due to contemporaneous failures of surrogate endpoints in cardiology trials, researchers were wary when they did not understand the role a surrogate played in disease pathogenesis and progression. They noted that the role and importance of a biomarker may change over the course of a disease, such that extension of results in a population with more advanced disease may not translate to a population with less advanced disease and vice

versa. Finally, researchers were not confident in the analytical validation of the tests being used to measure the surrogate endpoints (Cotton, 1991). In 1992, Lagakos and Hoth noted that experience from use of CD4 cell count as a surrogate endpoint in HIV/AIDS trials led to the idea that "it seems unrealistic to expect that any single marker can fully explain all of a drug's clinical effects." Furthermore, they recommended that "we cannot confidently abandon clinical endpoints as the basis for judging efficacy in these large trials. . . . It is therefore important that we continue to conduct comparative efficacy trials that collect data on both clinical outcomes and surrogate markers to establish CD4 count or other markers as valid surrogates for clinical effect" (Lagakos and Hoth, 1992). In 1993, Holden noted the desire of some to obtain a list of preapproved surrogate endpoints has been worrying to regulators because of the relevance of a biomarker's context of use in every application. In the article, Holden summarized a statement of Sidney Wolfe of the Public Citizen Health Research Group, saying that "drug companies could abuse [approvals of surrogate endpoints by the FDA] by failing to do careful clinical trials once they get a marker approved. . . . If clinical trials don't pan out, it might be very hard to ban the unapproved drug" that had been provisionally approved on the basis of the proposed surrogate endpoint (Holden, 1993).

Several of these warnings have been repeated since the early 1990s. Psaty et al. (1996) pointed out that different blood pressure-lowering interventions do not result in the same effects on clinical outcomes for a given reduction in blood pressure. De Gruttola et al. (1997) noted that unless disease mechanism of action is understood, uncertainty is inherent in the assumption that the surrogate can predict all of an intervention's effect. Schatzkin and Gail (2002) discussed use of surrogate endpoints in cancer research in 2002; they again noted the difficult balance between strong evidence that a surrogate endpoint has predictive value for the clinical endpoint and use of surrogates to achieve new drug approvals before full clinical trials using clinical endpoints can be completed. In the same year, DeMets and Califf (2002) reviewed principles of cardiovascular research and focused on the important distinctions between putative surrogate endpoints and clinical endpoints, reviewing multiple cases in which naïve use of putative surrogates had endangered patients with cardiovascular disease. In these cases, therapies, including antiarrhythmic, heart failure, and antiatherosclerosis treatments that had been assumed to be beneficial based on putative surrogate endpoints were indeed detrimental to health when confirmatory trials were done, usually because of off-target effects of systemically administered drugs. Manns et al. (2006) cited problems with the use of surrogate endpoints in a 2006 editorial. They discussed the opportunity cost of making decisions about allocation of healthcare resources (monetary, professional, and tangible),

treatment decisions to use one treatment and forgo others, and allocation of research funding. The authors suggest that "it would seem prudent for [clinical practice guideline] developers to refrain from recommending the use of new agents until they have been proved to improve clinically meaningful outcomes" (Manns et al., 2006). Krumholz and Lee (2008) wrote in the *New England Journal of Medicine* that although use of surrogate endpoints can simplify the practice of medicine, it can do so at the cost of quality and outcomes. In 2009, Colatsky noted that surrogate endpoint biomarkers, low-density lipoprotein cholesterol (LDL-C) levels and carotid intima-media thickness (IMT) in this example, do not always correlate well with one another, making interpretation of trial results difficult (Colatsky, 2009).

These cautionary statements have gathered strength as some surrogate endpoints have failed. Examples of these failures and the reasons for their occurrence are discussed in the next section.

Failure of Surrogate Endpoints: Reasons and Examples

Putative surrogate endpoints often fail to predict clinical outcomes. In 1996, Fleming and DeMets published a paper explaining the failures in surrogate endpoints that had occurred mostly during the late 1980s and early 1990s (Fleming and DeMets, 1996). As described in Figure 2-1, according to Fleming and DeMets (1996), several factors explain the failure of surrogate endpoints: (1) the surrogate endpoint does not involve the same pathophysiologic process that results in the clinical outcome; (2) the intervention affects only one pathway mediated through the surrogate, of several possible causal pathways of the disease; (3) the surrogate is not part of the causal pathway of the intervention's effect, or is insensitive to its effect; and (4) the intervention has mechanisms of action independent of the disease process. As noted in Figure 2-2, the most promising setting in which to qualify a surrogate endpoint occurs when the surrogate is on the only causal pathway of the disease process, and the intervention's entire effect on the clinical outcome is mediated through its effect on the surrogate (Fleming and DeMets, 1996). However, even in the best of circumstances, it is possible for surrogate endpoints to be misleading by either overestimating or underestimating an intervention's effect on clinical outcomes.

A number of biomarkers have been proposed as rational surrogate endpoints, but have failed to demonstrate usefulness for that purpose upon further scrutiny in clinical trials. One example was the use of beta-carotene and retinol as biomarkers for cancer, cardiovascular disease, and (later) cataract risk, and as interventions for chemoprevention of these diseases. Observational studies indicated that lower dietary intakes of

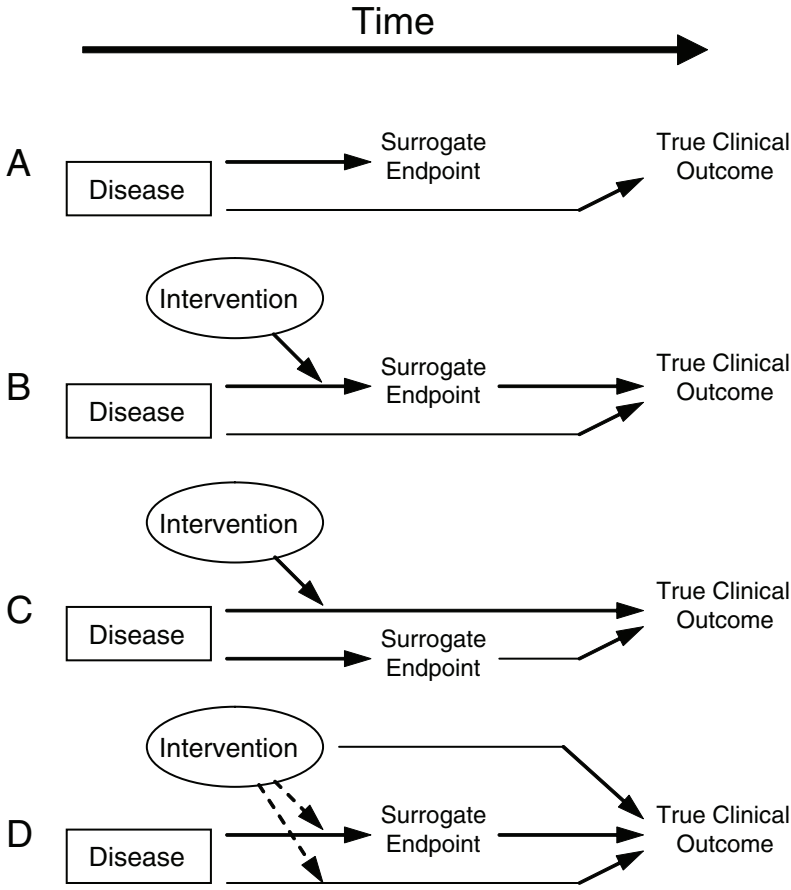

FIGURE 2-1 Reasons for failure of surrogate endpoints. (A) The surrogate is not in the causal pathway of the disease process. (B) Of several causal pathways of disease, the intervention affects only the pathway mediated through the surrogate. (C) The surrogate is not in the pathway of the intervention's effect or is insensitive to its effect. (D) The intervention has mechanisms of action independent of the disease process. Dotted lines = mechanisms of action that might exist.
SOURCE: Fleming and DeMets (1996). Reprinted, with permission, from the *Annals of Internal Medicine*. Copyright 1996 by American College of Physicians.

beta-carotene and lower serum levels of beta-carotene were associated with greater risk of cancer. It is useful to note that while serum level of beta-carotene is a biomarker for adequate intake of the nutrient and a proposed surrogate endpoint for prevention of cancer and atherosclerotic disease, supplementation of the diet with beta-carotene is an intervention to either address deficiencies or conditions for which it is used as a sur-

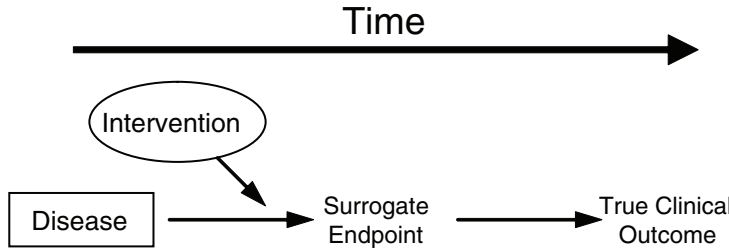

FIGURE 2-2 The setting that provides the greatest potential for the surrogate endpoint to be valid.
SOURCE: Fleming and DeMets (1996). Reprinted, with permission, from the *Annals of Internal Medicine*. Copyright 1996 by American College of Physicians.

rogate. Beta-carotene was shown to have in vitro antioxidant effects, and supplementing the diet with beta-carotene as a dietary supplement was expected to lower risk for atherosclerotic disease and cancer. However, its use in large population studies with mortality as the endpoint was not shown to lower risk for atherosclerosis or cancer; instead, it was shown to increase cancer incidence (Omenn et al., 1996; Peto et al., 1981). Beta-carotene will be discussed further in Chapter 4.

In another example, elevated serum levels of homocysteine were found to be associated with greater risk for atherosclerotic disease in observational associations and serum homocysteine was thought to be a surrogate endpoint. Homocysteine can exacerbate endothelial dysfunction, thrombosis, and other risk mechanisms for atherosclerosis. Folic acid was shown to decrease levels of circulating homocysteine. Researchers were confident that cardiovascular endpoints of death and vascular morbidity would be reduced with the administration of folic acid supplements. During this period, the use of folic acid supplements was found to decrease fetal development of neural tube defects when administered to pregnant women, and grain products were fortified with folic acid in the United States and other countries. The incidence of neural tube defects decreased following fortification. However, atherosclerotic disease, either coronary heart disease or peripheral vascular disease, did not decrease following folic acid fortification or with the administration of folic acid supplements in several large clinical trials despite important decreases in serum homocysteine levels with both interventions (Clarke et al., 2007).

From these examples, it is apparent that without a detailed understanding of a biomarker's role in the disease or treatment mechanism, biomarker evaluation can be difficult. The recent failure of some sur-

rogate endpoints to predict clinical outcomes has elicited concern over guidelines and performance measures used in clinical decision making. Traditionally, clinicians focus on reducing risk factors below certain levels to prevent disease; for example, clinical guidelines and performance measures "encourage treatment geared toward achieving ambitious goals for levels of glycated hemoglobin, lipids, and blood pressure" (Krumholz and Lee, 2008). In light of recent trials that demonstrate a reduction in a risk biomarker without a corresponding reduction in risk, Krumholz and Lee suggest a rethinking of risk factor reduction. Instead of focusing on just the amount a risk biomarker is reduced, clinicians should also be aware of the strategy involved in risk reduction. According to Krumholz and Lee (2008), "We are now beginning to appreciate that a strategy's effect on a risk biomarker may not predict its effect on patient outcomes." Since it is recognized that "[s]ome strategies are known to improve patient outcomes, whereas others are known to affect only risk-factor levels or other intermediate outcomes," Krumholz and Lee believe that guidelines and performance measures should not specify targets without strategies used to achieve them. Additionally, practice guidelines and performance measurement should discuss risks of disease and adverse events in a more sophisticated and explicit way so that an assessment of net clinical benefit can be made (Krumholz and Lee, 2008).

As Krumholz and Lee (2008) pointed out, changes in surrogate endpoints do not always correspond with changes in clinical outcomes. Data from additional clinical trials have supplemented the notion that effects on proposed surrogate endpoints may fail to predict clinical outcomes. Nambi and Ballantyne (2007) emphasized that "we must use a great deal of caution before substituting a surrogate for a clinical endpoint" because the scientific community has been misled by biomarkers in the past. Patients and the credibility of science in the eyes of the public can be negatively impacted when the scientific community is misled by a biomarker. Fleming and DeMets (1996) further noted that "a review of recent experiences with surrogates is sobering, revealing many cases for which biological markers were correlates of clinical outcomes but failed to predict the effect of treatment on the clinical outcome." The following examples related to cardiovascular disease (CVD)—arrhythmia suppression, exercise tolerance in congestive heart failure, and lowering lipids—were outlined by Fleming and DeMets as telling examples of failed surrogate endpoints.

Arrhythmia Suppression

As described by Fleming and DeMets (1996), an example of the failure of a surrogate endpoint to predict clinical outcomes is the reduction

of ventricular ectopic contractions for decreased cardiovascular mortality. When drugs were being developed and clinically tested, it was well known that compared to patients without ventricular arrhythmia, ventricular arrhythmia was independently associated with a significant increase in the risk of death related to cardiac complications, including sudden death (Bigger et al., 1984; Cardiac Arrhythmia Suppression Trial [CAST] Investigators, 1989; Echt et al., 1991; Mukharji et al., 1984; Ruberman et al., 1977). Researchers hypothesized that suppression of ventricular arrhythmias after myocardial infarction would reduce the rate of death. Scientists were so confident in this hypothesis that three drugs were approved by the FDA—encainide, flecainide, and moricizine—using arrhythmia suppression as the surrogate endpoint in phase III clinical trials. To illustrate the confidence scientists had in arrhythmia suppression as a surrogate endpoint, many of them believed that randomizing patients to either one of the study drugs or a placebo would be unethical. After approvals based on positive echocardiogram data, a feasibility trial was first conducted to determine whether a placebo-controlled trial would be safe enough to undertake (Cardiac Arrhythmia Pilot Study [CAPS] Investigators, 1986, 1988; CAST Investigators, 1989; Emanuel and Miller, 2001; Ruskin, 1989). After approval, more than 200,000 people eventually took these drugs each year, despite the lack of data evaluating the reduction of arrhythmias on mortality rates. The Cardiac Arrhythmia Suppression Trial (CAST) was designed to assess the drugs' impact on survival for patients who had had myocardial infarction and at least 10 premature beats per hour. Both the encainide and flecainide arms of the trial were terminated early when 33 sudden deaths occurred, as compared to only 9 in the matching placebo group. In total, 56 patients in the encainide and flecainide groups died, compared to 22 patients in the placebo group. Later data confirmed that patients taking moricizine were also at increased risk for death (Fleming and DeMets, 1996).

In addition to the CAST study, two other examples of failed surrogate endpoints have occurred with arrhythmia treatment. Quinidine had been used for many years to restore and maintain sinus rhythm in patients with atrial fibrillation. However, a meta-analysis indicated that quinidine increased the mortality rate from 0.8 percent to 2.9 percent, which outweighed the benefit of maintaining sinus rhythm (Fleming and DeMets, 1996). According to Lesko and Atkinson (2001), "unanticipated adverse consequences of drug therapy are a frequent confounding factor when biomarkers [such as maintaining normal sinus rhythm] are relied on as surrogates for definitive endpoints." Ventricular tachycardia, in the case of lidocaine drug therapy, was also shown to be an inadequate surrogate endpoint. Although a meta-analysis indicated lidocaine therapy produced a one-third reduction in the risk of ventricular tachycardia, it was also

accompanied by a one-third increase in death rate (Fleming and DeMets, 1996). The failure of surrogate endpoints (e.g., maintenance of normal sinus rhythm and reduction of risk of ventricular tachycardia) to predict clinical endpoints "underlies much of the controversy surrounding the use of surrogate endpoints as the basis for regulatory evaluation of new therapeutic entities" (Lesko and Atkinson, 2001).

Exercise Tolerance in Congestive Heart Failure

Decreased cardiac output, decreased exercise capacity, and high risk of death are conditions associated with congestive heart failure, noted Fleming and DeMets (1996). Heart failure is a leading problem in cardiology; for example, 12 percent of a cohort of individuals age 65 or over were found to have symptomatic heart failure (Afzal et al., 2007). Heart failure patients may experience shortness of breath, congestion in the lungs, difficulty exercising, swelling in the legs, and quality-of-life–reducing effects. During the time leading up to the Prospective Milrinone Survival Evaluation (PROMISE) trial, cardiac output and ejection fraction had been used as surrogate endpoints, while exercise tolerance and symptomatic improvement had been used as intermediate endpoints. The PROMISE trial was requested by the FDA, which was concerned about long-term adverse effects of milrinone (Fleming and DeMets, 1996). Milrinone, a drug that was used to treat congestive heart failure, was shown to increase total mortality in the PROMISE trial, even though earlier studies demonstrated milrinone's effectiveness in improving cardiac output and increasing exercise tolerance. The drug flosequinan, a vasodilator that reduces cardiac workload, was also conditionally approved by the FDA to treat congestive heart failure in patients who did not respond to or tolerate other drugs. However, the Prospective Flosequinan Longevity Evaluation (PROFILE) trial demonstrated that flosequinan increased total mortality, even though it improved exercise tolerance. According to Fleming and DeMets (1996), "[a]lthough cardiac output, ejection fraction, and exercise tolerance are correlated with longer survival of patients with congestive heart failure, a treatment-induced improvement in those measurements is not a reliable predictor of the effect of treatment on mortality rates."

EVALUATION FRAMEWORKS

Biomarkers differ in their contexts of use and thus in the types of evidence needed for evaluation. Furthermore, use of surrogate endpoints for collection of evidence in support of policy or regulatory decisions is subject to the challenges and risks discussed in the previous sections (see Figures 2-1 and 2-2 and associated discussion). For additional detail, see Figure 2 in the paper by Boissel et al. (1992), outlining an approach for

selection of surrogate endpoints. As each of these figures illustrate, the evaluation of a biomarker as a surrogate endpoint is particularly challenging because of the biological complexity of human disease and response to drugs and nutrients. Neither correlation of the biomarker with clinical outcome nor biological plausibility is sufficient to establish the usefulness of a biomarker as a surrogate endpoint. Moreover, qualification of a biomarker for a particular disease or treatment does not necessarily translate to qualification for related uses or even for an essentially identical use at a different point of time (and thus a different context of use).

Several frameworks for biomarker qualification and several for biomarker assay validation have been published. Appendix A presents a time line of critical developments in the discussion about biomarker and surrogate endpoints evaluation, republished with permission from the 2008 review in *Statistical Methods in Medicine* by Lassere. Terminology is presented as it was by Lassere, which was consistent with the original publications. Since 2007, there have been a few important publications, which have also been tabulated in Appendix A.

The next section discusses the evolution of thought on association and causation between exposure to a pathogenic agent, biomarkers, and incidence and mortality from disease. Several examples of the evaluation and use of surrogate endpoints in drug development are then discussed. The last two sections address the two main directions in the discussion of biomarker evaluation: those focusing on statistical methods and those focusing on qualitative methods. The reason is that while it is straightforward to establish a statistical association, it is difficult to definitively establish causality. Qualitative criteria have been used to fill this gap in the quantitative methods. Furthermore, decisions sometimes must be made when sufficient data are not available to make a quantitative analysis, and so qualitative methods are used.

Biomarker–Clinical Endpoint Relationships: Association Versus Causation

Many students of biology and epidemiology are familiar with Koch's postulates for determining the cause of infectious diseases. These postulates state that in order to conclude that a particular infectious agent is the cause of a disease, the following conditions must be fulfilled:

1. The agent must be associated with all cases of the disease;
2. The agent must be isolable and cultured from a diseased organism;
3. The cultured agent must be able to infect a new host; and
4. The agent must be reisolable from the host in postulate 3.

These postulates were developed in the 1880s, and in the 1900s, scientists sought to establish causality in diseases that were not infectious, such as cancer. In a report outlining the evidence supporting a causal link between smoking and lung cancer, an advisory committee to the Surgeon General of the Public Health Service outlined five criteria for the case of non-infectious or chronic diseases: the strength, specificity, temporality, and consistency of the association (Advisory Committee to the Surgeon General, 1964). These criteria were refined when, in 1965, Sir Austin Bradford Hill discussed these criteria in a famous lecture to the section of occupational medicine of the UK's Royal Society of Medicine (Hill, 1965). The criteria are now known as Hill's criteria and are outlined in Box 2-2. Since the 1960s, these criteria have been used in environmental health, toxicology, pharmacology, epidemiology, and medicine.

Surrogate endpoints have been discussed for a little over 20 years. In 1989, Ross Prentice defined the term "surrogate endpoint" in his paper entitled "Surrogate endpoints in clinical trials: Definition and operational criteria" (Prentice, 1989). This paper was accompanied by three other papers in an issue of *Statistics in Medicine* exploring the possible use of biomarkers as surrogate endpoints, using examples from cancer (Ellenberg and Hamilton, 1989), cardiovascular disease (Wittes et al., 1989), and ophthalmologic disorders (Hillis and Seigel, 1989). As discussed briefly in the previous chapter, the Prentice criteria specify that a biomarker under consideration as a potential surrogate endpoint must correlate with the clinical outcome it is meant to replace and that the biomarker must capture the entire effect of the intervention on the clinical endpoint (Prentice, 1989). Further development of statistical methods has occurred since 1989, as statisticians search for methods to ease the burden of the second criterion (Fleming, 2005). These approaches include meta-analysis of data from multiple trials (Alonso et al., 2006; Burzykowski et al., 2004; Buyse and Molenberghs, 1998; Buyse et al., 2000; Hughes, 2002; Hughes et al., 1995) as well as addressing the following: (1) the proportion of treatment effect described by the surrogate endpoint; (2) the relative effect and adjusted association; and (3) the surrogate threshold effect. These methods are summarized in Lassere's (2008) review, and several of them are discussed in detail in this chapter's section on statistical approaches to biomarker evaluation.

Nonetheless, surrogate endpoints were used before these conversations began. One of the best examples of this is blood pressure, which is used as a surrogate endpoint for CVD clinical outcomes. Blood pressure represents the historical course of biomarker evaluation, gradual accumulation of data, and agreement among stakeholders on the utility of a biomarker, as described in the earlier section on the history of the evaluation of blood pressure as a surrogate endpoint.

BOX 2-2
Hill's Criteria

1. **Strength—**Causation is supported if the relative risk due to the exposure is very large.
2. **Consistency—**Causation is supported if the relationship is seen in different populations at different times and in different circumstances.
3. **Specificity—**Causation is supported if an exposure appears to cause only a specific effect.
4. **Temporality—**Causation is supported if the exposure precedes the effect.
5. **Biological Gradient—**Causation is supported when the magnitude of the exposure is proportional to the magnitude of the effect.
6. **Plausibility—**Data elucidating the biological pathways leading from exposure to effect are useful.
7. **Coherence—**"The cause-and-effect interpretation of [the] data should not seriously conflict with the generally known facts of the natural history and biology of the disease."
8. **Experiment—**In some circumstances, evidence that removing the exposure lessens or removes the effect can be used to draw conclusions about causality.
9. **Analogy—**In some circumstances, comparison between weaker evidence of causation between an exposure and its effect and strong evidence of causality between another exposure and its similar effect is appropriate.

SOURCE: Hill (1965).

HIV/AIDS drug development provides another historical example of the use of surrogate endpoints. On October 11, 1988, frustrated with the length of time-to-approval for new therapies to treat HIV infection, ACT-UP, an AIDS patient advocacy group, staged a demonstration in front of FDA headquarters. Eight days later, on October 19, Frank Young, then commissioner of the FDA, announced regulations by which review times would be shortened for drugs designed to treat "life-threatening or severely debilitating" diseases (Arno and Feiden, 1988; AVERT, 2009; FDA, 1988). For that reason, HIV/AIDS drugs were some of the first to be approved explicitly on the basis of surrogate endpoints, and served as the foundation for the laws on accelerated approval of drugs and biologics. HIV/AIDS was also the first example of a more systematic, prospective approach to biomarker evaluation, although its precedent was not easily translatable into general guidance.

Finally, after the early 1990s, much of the literature has focused on the use of surrogate endpoints to approve oncology drugs. There is a substantial literature in this area, which is discussed in relation to use of

tumor size as a surrogate endpoint for cancer treatment interventions in Chapter 4. Research and development in oncology has been working to implement broader use of biomarkers, but this effort continues with lack of a standard approach.

Statistical Approaches to Biomarker Evaluation

Although randomized clinical trials with clinically meaningful endpoints provide the most rigorous means of assessing benefit of an intervention, such trials may be lengthy and expensive, and not always feasible. Therefore considerable interest has been shown in development of a framework for "statistical validation" of surrogates for clinical endpoints that can reliably provide information more quickly and cheaply about medical interventions. While much work has been done in this area, there remains no widely accepted research paradigm for statistical validation, in the way that, for example, randomized clinical trials provide such a paradigm for comparing new to existing therapies. Below we describe why no single paradigm is likely to arise soon, or perhaps ever. We also show, however, that existing frameworks and methods are useful for investigating the properties of surrogate endpoints.

It is useful to restate Prentice's influential definition of a statistically valid surrogate, which required that a test of the null hypothesis of no relationship of the surrogate endpoint to the treatment assignment must also be a statistically valid test of the corresponding null hypothesis based on the true endpoint (Prentice, 1989). Statistical validation was based on two conditions: (1) correlation of the surrogate with the true clinical endpoint; and (2) the ability of the surrogate to fully capture the treatment's "net effect" on the clinical endpoint. As described by Fleming and DeMets (1996), the net effect is the aggregate effect accounting for all mechanisms of action of the intervention. Considerable effort has been made to assess the degree to which this second condition holds in a variety of settings, but such analyses are complicated by difficulty in reliably estimating the quantities of interest and in the need for extensive assumptions (see below).

An alternative approach is based on meta-analyses across studies. Daniels and Hughes (1997) used Bayesian methods to construct prediction intervals for the true difference in clinical outcome associated with a given estimated treatment effect on the potential surrogate. By "borrowing" information regarding estimates of the effects of treatment on the clinical endpoint, and on the relationships between the surrogate and the clinical endpoint given treatment from previous studies, one predicts effects of a new treatment from data on the surrogate.

An important recent paper by Joffe and Greene (2009) attempts to pro-

vide a broader intellectual framework, using ideas from causal inference, that subsumes several different approaches (including those described above) and also provides insight into why this research is so challenging. They describe four different frameworks for statistical validation of surrogacy, and show connections among them. The first is based on the Prentice criteria described above. A second considers the estimation of direct and indirect effects of treatment; the latter are those mediated through a biomarker. Joffe and Greene describe these two approaches as belonging to a category of causal effects frameworks, in which knowledge of the effects of the treatment on the surrogate and of the surrogate on the clinical outcome is used to predict the effect of the treatment on the clinical outcome.

The use of causal graphs modeling shows the challenge of basing a statistical validation procedure on the Prentice criteria. For true surrogate markers, there should be no direct effect of treatment independent of the marker, but instead all of the effect should be mediated by the surrogate. If there were no other causes of the clinical endpoint besides the treatment and the surrogate, analyses would be straightforward; in reality many other factors are likely to be involved. While randomization assures that treatment is not associated causally with any confounding variable, there is no reason to believe this to be true for the surrogate. In fact, the relationship of surrogate to clinical endpoint may well be confounded by other variables, each of which may or may not be measured. Joffe and Greene point out that even if the surrogate mediates the entire effect of treatment on the outcome (a most unlikely situation), the presence of confounding factors would imply that the treatment is not independent of the endpoint given the surrogate—in other words the Prentice criteria will not be met.

Model-based estimation of direct and indirect effects, possibly making use of the causal modeling approaches of Robins and Greenland (1992, 1994), offer some hope of addressing this issue, but such methods still require strong assumptions. One such assumption is that the intervention directly affects the surrogate, which in turn affects the clinical endpoint. Another is that one can control for confounding of the effect of the surrogate on the clinical outcome by proper inclusion of baseline covariates in a regression. In reality, baseline covariates may not be sufficient—an occasion that arises when a postrandomization covariate, influenced by treatment, affects the surrogate and is independently associated with outcome. For example, suppose that a blood pressure medication induced fatigue and therefore caused a reduction in the amount of exercise patients undertook; such an adverse consequence of treatment could affect both blood pressure and clinical events, such as time to myocardial infarction. Procedures are available to permit assessment of surrogacy in this situation, but

they require that the confounding be controllable, through measurement and appropriate modeling. Unfortunately, there can be no way to test such that confounding can be appropriately controlled.

The third framework mentioned by Joffe and Greene is that of meta-analysis. As described above, meta-analysis investigates the relationship of the effects of treatment on surrogates with its effects on clinical outcomes over a series of trials. The fourth framework is defined in terms of the ideas of principal stratification, developed by Frangakis and Rubin (2002). These approaches belong to the causal-association paradigm, in which the effect of treatment on the surrogate is associated, across studies or population groups, with its effect on the clinical outcome, thereby allowing prediction of the effect on the clinical outcome from the effect on the surrogate.

For the meta-analysis approach, the average value of the surrogate measured in each trial should be able to predict the outcome for that trial. Of course, such an approach requires variability in the effect of treatment on the surrogate across studies. This approach may be the most promising because of its avoidance of the need for strong assumptions regarding confounding; nonetheless, even in this case, interpretation must be made with care. For example, Daniels and Hughes (1997) demonstrated that the change in CD4 count was associated with clinical endpoints (time to new AIDS definition or death). But in their example, all of the studies with large treatment and surrogate effects compared active treatments to placebo, whereas all of the studies with small treatment or surrogate effects had active controls. Therefore, extension of the results to a setting where a trial with an active control had a strong surrogate effect may not be warranted, as the biological processes might be quite different in this case than among those that were studied.

In contrast to the meta-analytic approaches, the principal surrogacy approach focuses on the association of the individual-level effects on surrogate and outcome. As is true in general of principal stratification, the group for whom the causal effects of treatment are defined is not observable, because for each individual, the surrogate can be observed only on one treatment and not the other(s). Full description of this approach is beyond the scope of this chapter, but such analyses are most likely to be useful in settings whether there is a strong effect of treatment on both surrogate and endpoint.

In conclusion, no simple paradigm for evaluation of surrogates is possible; consistency of findings across all of the approaches described by Joffe and Greene would probably provide the most convincing evidence. But the statistical methods do not in themselves provide the type of compelling evidence that a randomized trial with nearly complete follow-up can provide. Both a deep understanding of biological context combined

with a thorough knowledge of causal research are necessary for any attempt at statistical validation of markers.

Decision Analysis Approaches to Biomarker Evaluation

Decision theory allows for logical and reproducible decision-making based on both quantitative and qualitative inputs. For biomarker evaluation, decision theory may be useful for the utilization step, and many principles from decision theory can be found throughout the report. Dr. Rebecca Miksad from Harvard University gave a presentation to the committee on decision theory as it could be applied to biomarker evaluation at the committee's April 2009 workshop. In the presentation, Miksad defined decision science as a "field of science which rigorously and quantitatively evaluates the short and long term outcomes of complex clinical situations through analysis of clinical decisions" (Miksad, 2009). Decision analysis formalizes complex decision-making processes involving ambiguity in data, variation in data interpretation, competing benefits and risks, gaps in information, and personal preferences when applicable. Decision analysis requires that decision makers break down decisions into their component parts and make any assumptions explicit. Miksad identified five unique features of decision analysis in her presentation (Box 2-3).

While analytical sensitivity and specificity of biomarker tests are important aspects of analytical validation, it is also important to take variability between individual interpreters of data. Receiver operating characteristic (ROC) graphs are a common decision analysis tool for accomplishing this goal. An ROC graph plots the impact of data interpretation variability on use of a given decision threshold, such as a cutoff value for a diagnostic test, for example (IOM, 2005). The x-axis of an ROC curve is the likelihood of a false positive result, or 1-specificity, while the y-axis of an ROC curve is the likelihood of a false negative result, or the sensitivity (IOM, 2005). ROC curves are described in Figure 2-3.

During decision analysis, all possible choices are mapped onto a decision tree. Then, mathematical models are used to compare possible outcomes of each choice. From these models, decision makers can then choose the most appropriate course of action or identify areas where more information is needed.

Miksad outlined important questions that can be addressed using decision analysis for biomarker evaluation (Miksad, 2009):

- What are the optimal characteristics and analytical thresholds for the biomarker assays themselves?
- What are the positive and negative predictive values of the biomarker assays?

BOX 2-3
Five Unique Features of Decision Analysis for Surrogate Endpoint Evaluation

- Directly addresses clinical complexity:
  - Multiple and potentially contradictory data
  - Multiple treatment options
  - Multiple potential interactions
  - Competing risks from patient comorbidities
- Explicitly incorporates uncertainty:
  - Data errors
  - Ambiguity and variations in data interpretation
  - Discordance between data and true disease state
  - Variable treatment effects, side effects and disease courses
- Identifies and compares trade-offs between competing objectives and risks:
  - Benefit of diagnosis versus risks of procedure
  - Therapeutic effects versus side effects
- Extends existing trial data to project outcomes across long time periods, including estimations of uncertainty
- Component parts of clinical decisions are broken down and data is recombined systematically

SOURCE: Miksad Presentation (2009). Reprinted, with permission, from Rebecca Miksad. Copyright 2009 by Rebecca Miksad.

- Does use of the biomarker assay lead to improved clinical outcomes?
- What are the areas of uncertainty that lead to the largest differences in predicted affects on clinical outcomes?
- Is additional data needed before use of the biomarker can be adopted?

Decision theory can be useful as a way to formalize the biomarker evaluation framework. While each biomarker evaluation would require a unique decision analysis, these analyses would provide stakeholders with a transparent accounting of the assumptions and subjective judgments that were needed for making specific decisions. In addition, these analyses would provide details on where biomarkers may benefit from the collection of additional data.

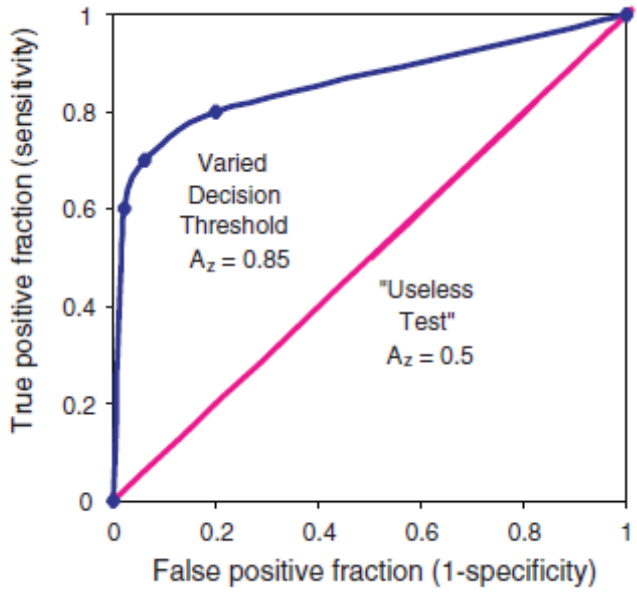

FIGURE 2-3 Receiver operating characteristic (ROC) graph of a varying decision threshold compared with a "useless test." The best-fit curve drawn through these points is the ROC curve, which represents the overall performance of the diagnostic test across all possible interpretations (decision thresholds). The overall accuracy of this test under varying conditions is determined by the area under the complete curve, 0.85. The leftmost point shows low sensitivity and high specificity. The middle point shows moderate sensitivity and specificity. The rightmost point shows high sensitivity and low specificity. Yet because they all lie on the same curve they have the same overall statistical accuracy, which is quantified by A_Z. The 45-degree-angle line represents a series of guesses between two choices, as in a coin toss. This would be considered a "useless test" if the outcome of the test was dichotomous (for example cancer vs. no cancer) for diagnostic purposes. For instance, radiologists reading mammograms with their eyes closed would tend to fall on this line. The number of true positives would approach the number of false negatives.The area under such a curve, 0.5, represents 50 percent accuracy of the test. In contrast, the ROC curve for a test with 100 percent accuracy will trace the y-axis up at a false-positive fraction of zero and follow along the top of the graph at a true-positive fraction of one. The area under such a curve would be 1.0 and represent a perfect test.
SOURCE: IOM (2005).

Qualitative Approaches to Biomarker Evaluation—Drug Development

This section describes one of the biomarker evaluation frameworks presented in the tables in Appendix A. In particular, this section discusses

efforts made through public–private partnerships to develop a standardized, fit-for-purpose biomarker evaluation process. Beginning in the late 1990s and early 2000s, drug developers began participating in the development of biomarker evaluation processes (Colburn, 1997, 2000; Wagner, 2002). This effort was further strengthened by the formation of public–private partnerships such as the Biomarker Consortium and other Foundation for National Institutes of Health (NIH) efforts, as well as the Critical Path Institute (C-Path). The frameworks proposed in collaborations with pharmaceutical industry representatives strive for several characteristics: reproducibility, clear process, risk management, and incremental or fit-for-purpose evaluations (Altar, 2008; Altar et al., 2008; Lathia et al., 2009; Wagner, 2002, 2008; Wagner et al., 2007; Williams et al., 2006). In addition, several also consider cost effectiveness in frameworks to make decisions on biomarker evaluation (Altar et al., 2008).

A 2008 paper proposed use of an "evidence map" for use in biomarker evaluations (Table 2-2) (Altar et al., 2008). This map was developed as a collaboration between pharmaceutical industry representatives, a representative from the Foundation for NIH, and an FDA representative. The paper subsequently received attention from FDA staff at a conference entitled "2008 Cardiovascular Biomarkers and Surrogate Endpoints Symposium: Building a Framework for Biomarker Application." Briefly, the evaluation method proposed involves use of a committee to make decisions based on data and non-quantitative factors, such as public tolerability of the proposed decision. The first step in the process is for the committee to define and agree on a purpose and context of use for the biomarker. The next step is to assess the potential benefits and harms of the future success or failure of the biomarker in its proposed use. The third step is to come to an agreement about the tolerability for risk for the particular biomarker, given its proposed purpose and context of use. The fourth step is to assess the evidentiary status of the biomarker through use of the evidence map. During this step, the purpose and context-of-use combination is given a grade the biomarker needs to achieve in order to be deemed qualified. The final step is to summarize the committee's proceedings for the stakeholders.

The authors of the paper tested this framework with a panel of experts at a workshop, and found it to be useful; they also suggested next steps to improve the framework (Altar et al., 2008). This framework provided some of the basis for Recommendations 1 and 2.

In 2009, many industry authors of the Altar et al. (2008) paper published a paper commenting on the use of surrogate endpoints for drug approvals. They described characteristics of successful surrogate endpoints: biologic plausibility, prognostic value, and a positive correlation

between an intervention's effect on the surrogate endpoint and the clinical endpoint (Lathia et al., 2009). A representative from CDER commented on their paper in the same issue, providing important examples of how biomarkers can be used to speed drug development without being used as surrogate endpoints (Gobburu, 2009).

Inclusion of Cost-Effectiveness Analysis in Biomarker Evaluation

A controversial issue in the drug development community is whether or not cost–benefit analysis should be part of a biomarker evaluation process. In 2006, Williams and colleagues outlined principles for biomarker evaluation that were the basis for the 2008 evidence map discussed above (Williams et al., 2006). Principle 8 was that "post hoc review of cost effectiveness should be performed at regular intervals as new information is available and conclusions recorded systematically as to how this should modify the qualification and use" (Williams et al., 2006). In 2008, this idea was discussed again: "some individuals from industry expressed great concern about the use and potential misuse of cost–benefit analyses and principles and did not wish to see them used here" (Altar et al., 2008).

Some additional considerations of the committee considered during its deliberations included the following:

- The FDA does not include analysis of cost in decisions to approve drugs or in other regulatory decisions.
- In their 2009 study entitled *The use of surrogate outcomes in model-based cost-effectiveness analyses: a survey of UK Health Technology Assessment reports*, Taylor and Elston stated that their "literature searches found no empirical studies examining the use of surrogate outcomes in [health technology assessments] and [cost-effectiveness models] therein."
- Conclusions regarding the cost-effectiveness analysis on drug development processes cannot be definitively drawn until evidence relating the use of a new intervention with clinical outcomes is available.

An explanation of why the committee did not include cost-effectiveness analysis as part of its biomarker evaluation process is included in Chapter 3.

TABLE 2-2 Altar et al. (2008) Proposed Evidence Map for Biomarker Qualification

Evidence Type	Grade D	Grade D+/C-	Grade C
Theory on biological plausibility	Observed association only	Theory, indirect evidence of relevance of the biomarker from animals	As for lower grade but evidence is direct
Interaction with pharmacologic target	Biomarker identifies target in in vitro binding		
Pharmacologic mechanistic response	In vitro evidence that the drug affects the biomarker	In vitro evidence that multiple members of this drug class affects the biomarker	In vivo evidence that this drug affects biomarker in animals
Linkage to clinical outcome of a disease or toxicity		Biomarker epidemiologically associated with outcome without any intervention	Biomarker associated with change in outcome from intervention in another drug class
Mathematics replication, confirmation		An algorithm is required to interpret the biomarker and was developed from the dataset	
Accuracy and precision (analytic validation)			
Relative performance		Does not meet performance benchmark	

SOURCE: Altar et al. (2008). Adapted, with permission, from Macmillian Publishers Ltd: Clinical Pharmacology and Therapeutics. Altar, C. A., D. Amakye, D. Buonos, J. Bloom, G. Clack, R. Dean, V. Devanarayan, D. Fu, S. Furlong, C. Girman, L. Hinman, C. Lathia,

Grade C+/B-	Grade B	Grade B+/A-	Grade A
Theory, indirect evidence of relevance in humans	Theory, direct evidence in humans, non-causal pathway possible	As for lower grade, but biomarker on causal path	Human evidence based on mathematical model of biology showing biomarker is on causal pathway
Biomarker identifies target in in vivo binding in animals	Biomarker identifies target in in vivo studies or from human tissue, no truth standard		Biomarker identifies target in in vivo studies or from tissues in humans, with accepted truth standard
As for lower grade but effect shown across drug class	Human evidence that this drug affects the biomarker OR animal evidence of specificity	Human evidence across this mechanistic drug class	Human evidence that multiple members of this drug class affect the biomarker and the effect is specific to this class/ mechanism
As for lower grade but in this drug class	As for lower grade but multiple drug classes albeit inconsistent or a minority of disease effect		As for lower grade but consistent linkage and explains majority of disease effect
Algorithm was developed from a different dataset and applied here prospectively			Algorithm developed from different dataset, replicated prospectively in other sets and applied prospectively here
Sources of technical variation are unknown but steps are taken to ensure consistent test application	Major sources of variation known and controlled to be less than biological signal; standardization methods applied		All major sources of technical imprecision are known, and controlled test/assay accuracy is defined against standards
Similar performance to benchmark			Exceed performance of benchmark or best alternative biomarker

L. Lesko, S. Madani, J. Mayne, J. Meyer, D. Raunig, P. Sager, S. A. Williams, P. Wong, and K. Zerba. 2008. A prototypical process for creating evidentiary standards for biomarkers and diagnostics. *Clinical Pharmacology and Therapeutics* 83(2):368–371, Copyright 2007.

EVOLUTION OF REGULATORY PERSPECTIVES ON SURROGATE ENDPOINTS

Table 2-3 outlines the regulations and guidances pertaining to surrogate endpoints and the FDA. FDA regulatory authority for drugs, biologics, devices, foods, and supplements is discussed in detail in Chapter 5. While not discussed in detail, the NIH has historically played a vital role in the discovery, development, and regulatory perspective toward biomarkers; this is discussed briefly in Box 2-4.

2006–2008: FDA Pilot Process for Biomarker Qualification

Federico Goodsaid and Felix Frueh developed a biomarker qualification pilot process at the FDA, in collaboration with C-Path (Goodsaid, 2008a, 2008b; Goodsaid and Frueh, 2006, 2007a, 2007b; Goodsaid et al., 2008). The FDA pilot process for biomarker qualification was designed to qualify biomarkers incrementally, based on the data that are available for drug development or clinical applications. A biomarker would first be qualified in a narrow context of use, and then the context of use would be expanded as additional information became available. The

TABLE 2-3 List of Regulations and Guidances Pertaining to Surrogate Endpoints

Regulation or Guidance	Significance
21 C.F.R. 314.510	Accelerated approval: drugs. "Surrogate - Approval based on a surrogate endpoint or on an effect on a clinical endpoint other than survival or irreversible morbidity."[a]
21 C.F.R. 601.41	Accelerated approval: biologics. "Surrogate - Approval based on a surrogate endpoint or on an effect on a clinical endpoint other than survival or irreversible morbidity."[a]
Guidance for industry: Available therapy (FDA, 2004)	This guidance states that "the approval of one therapy under the accelerated approval regulations (either on the basis of a surrogate endpoint or with restricted distribution) should not preclude the approval under the accelerated approval regulations of additional therapies."

TABLE 2-3 Continued

Regulation or Guidance	Significance
21 C.F.R. 314.520	Postmarket authority of Food and Drug Administration (FDA) for drug accelerated approvals: "Restricted - Approval with restrictions to assure safe use."[a]
21 C.F.R. 601.42	Postmarket authority of FDA for biologic accelerated approvals: "Restricted - Approval with restrictions to assure safe use."[a]
21 C.F.R. parts 862–872, among others	The C.F.R. mentions surrogate endpoints in exceptions to the exemption of class I and II medical devices from premarket review: devices measuring analytes that are to serve as surrogate endpoints must undergo premarket review.
Guidance for industry and FDA staff: Postmarket surveillance under section 522 of the Federal Food, Drug, and Cosmetic Act (CDRH, 2006)	Postmarket surveillance may be requested when "premarket evaluation of the device may have been based on surrogate markers. Once the device is actually marketed, postmarket surveillance may be appropriate to assess the effectiveness of the device in detecting or treating the disease or condition, rather than the surrogate."
Guidance for industry: Clinical studies section of labeling for human prescription drug and biological products—Content and format (FDA, 2006a)	This guidance document recommends that manufacturers include more information in the Clinical Studies section of the label when "The study uses an unfamiliar endpoint (e.g., a novel surrogate endpoint), or there are important limitations and uncertainties associated with an endpoint."
Guidance for industry: Clinical data needed to support the licensure of seasonal inactivated influenza vaccines (CBER, 2007)	The document states that "For influenza vaccines, the immune response elicited following receipt of the vaccine may serve as a surrogate endpoint that is likely to predict clinical benefit, that is, prevention of influenza illness and its complications."
Guidance for industry: Clinical trial endpoints for the approval of cancer drugs and biologics (FDA, 2007)	The document describes current and past thought on use of non-survival endpoints in oncology approvals. A table comparing important cancer endpoints is presented.

continued

TABLE 2-3 Continued

Regulation or Guidance	Significance
Guidance for industry and FDA staff: Clinical study designs for catheter ablation devices for treatment of atrial flutter (CDRH, 2008)	The document states that "acute procedural success may be appropriate to serve as a surrogate effectiveness endpoint for catheters provided all of the following device characteristics are present: • Creates endocardial lesions • Manipulated in the endovascular space • A single ablation electrode • The energy source is radiofrequency (RF) • Temperature sensing capability • 'Steerable' (i.e., catheter has a tip which is manually-deflectable via a thumb-wheel or similar mechanism residing on the handle of the catheter) • Percutaneous placement."
Guidance for industry: Evidence-based review system for the scientific evaluation of health claims (CFSAN, 2009a)	Includes the definition of surrogate endpoint discussed in Chapter 1. The document lists the four currently accepted surrogate endpoints for health claims: "(1) serum low-density lipoprotein (LDL) cholesterol concentration, total serum cholesterol concentration, and blood pressure for cardiovascular disease; (2) bone mineral density for osteoporosis; (3) adenomatous colon polyps for colon cancer; and (4) elevated blood sugar concentrations and insulin resistance for type 2 diabetes." However, it also stipulates that biomarkers not on the biological pathway of a particular nutrient–disease risk link may not be used as surrogate endpoints for development of health claims.

NOTE: [a] http://www.fda.gov/Drugs/DevelopmentApprovalProcess/HowDrugsareDevelopedandApproved/DrugandBiologicApprovalReports/ucm121606.htm.

qualification process, as outlined in Figure 2-4, involves FDA reviewers, outside experts, and advisory committees. The process started with a two-page letter submitted to the FDA. The letter includes a description of the biomarker, an accurate definition of the context of use that the biomarker is being proposed for, and a list of the data supporting the request. Submissions are made by companies, consortia, and academics.

BOX 2-4
FDA's Risk Communication Advisory Committee

The FDA's Risk Communication Advisory Committee was created in 2008 with the following purpose:

> The Committee advises the Commissioner of the Food and Drugs or designee on methods to effectively communicate risk associated with products regulated by the Food and Drug Administration and in discharging responsibilities as they relate to helping to ensure safe and effective drugs for human use and any other product for which the Food and Drug Administration has regulatory responsibility. The Committee reviews and evaluates strategies and programs designed to communicate with the public about the risks and benefits of FDA-regulated products so as to facilitate optimal use of these products. It also reviews and evaluates research relevant to such communication to the public by both FDA and other entities, and facilities interactively sharing risk and benefit information with the public to enable people to make informed independent judgments about use of FDA-regulated products. (FDA, 2010)

The committee is currently chaired by Dr. Baruch Fischoff, professor in the Departments of Social & Decision Sciences and Engineering & Public Policy at Carnegie Mellon University. The committee has ten additional members. The committee meets four times a year. In 2009, the committee discussed topics such as

- Risk communication research needs,
- Quality of consumer drug information,
- Communicating about food recalls and food-borne illness,
- Communicating about tobacco and health,
- Clinical trials database, and
- Use of social media as surveillance tools.

SOURCE: FDA (2010).

The next step is the recruitment of a biomarker qualification review team. A briefing document is requested from the group submitting the request, and then a face-to-face meeting is held between the review team and the group submitting the request. The gaps in evidence are evaluated, revised data packages are requested, and the process goes back and forth until the package is as complete as possible. Then, the review team writes a document, and a regulatory briefing is submitted (Goodsaid et al., 2008). Goodsaid emphasized in his presentation at the Cardiovascular Markers of Disease (CMOD) conference that "*biomarker qualification* is the process by which data are provided to show that exploratory biomarkers are qualified for application in a specific context of use," and that "the *context of use* for a biomarker is the general area of biomarker application, specific

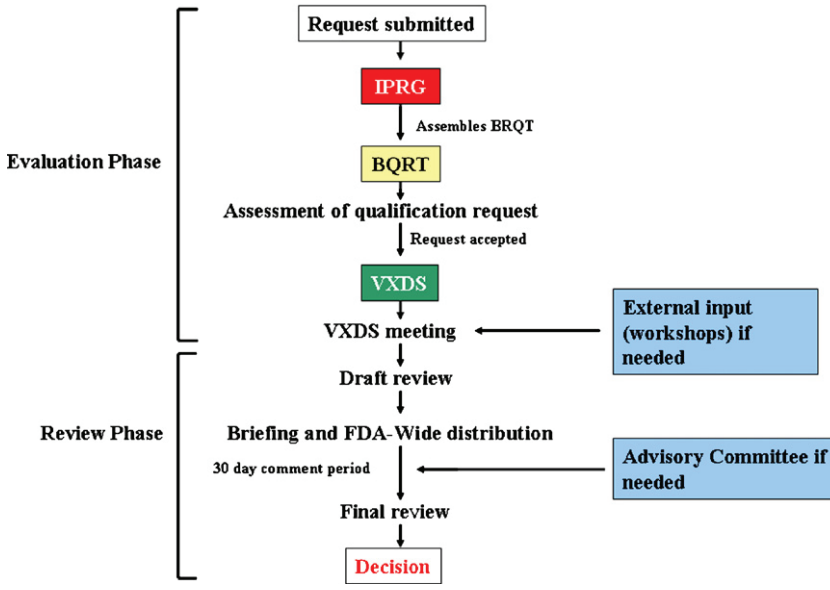

FIGURE 2-4 Outline of the Food and Drug Administration's (FDA's) biomarker qualification pilot process.
NOTE: BQRT = Biomarker Qualification Review Team; IPRG = Interdisciplinary Pharmacogenomic Review Group; VXDS = voluntary data submission.
SOURCE: Goodsaid et al. (2008). Reprinted with permission from Elsevier. Copyright 2008, Elsevier.

applications/implementations and critical factors which define where a biomarker is to be used and how the information from measurement of this biomarker is to be integrated in drug development and regulatory review" (Goodsaid, 2008a).

Melanie Blank, medical officer in the Division of Cardiovascular and Renal Products in CDER, has also discussed the FDA pilot process for biomarker qualification and how the evidentiary standards would be higher when the consequences of false results are graver (Blank, 2008): the qualification process as it would be applied to several problems such as how efficacy biomarkers can help in large, expensive drug trials where the clinical endpoint is rare and delayed, how safety biomarkers contribute when there is late discovery of toxicity resulting in late abandonment of the drug development program, and how safety biomarkers contribute when there are no sensitive methods to detect observed preclinical toxicities.

Example: Biomarkers of Kidney Toxicity

Dr. Joseph Bonventre of Harvard University spoke to the committee members at their first meeting. He has been involved on FDA committees as well as the only academic participant in the Critical Path Institute's biomarker qualification effort, which was done in collaboration with Federico Goodsaid at the FDA and industry partners. The Predictive Safety Testing Consortium (PSTC), as part of the Critical Path Institute's efforts in the area of biomarker evaluation, assembled a panel of scientists to evaluate potential safety biomarkers of acute kidney injury. These biomarkers are needed for use in "early diagnosis, to monitor severity and progression of disease, predict an outcome without an intervention, better stratify patients for clinical trials, predict who will respond to an intervention, [determine whether] the intervention [is] working ([through use of a] surrogate [endpoint]), and to identify therapeutic targets for an intervention" (Bonventre, 2009).

The most commonly used biomarkers for kidney injury are functional biomarkers rather than biomarkers of injury: serum creatinine and blood urea nitrogen. As in many organ systems, there are different stages of injury: risk, damage, reduction in function, organ failure, and death. Complications are associated with each stage. Elevations of serum creatinine and blood urea nitrogen above established normal ranges occur only after significant renal damage is present. Biomarkers of injury were the target of the preclinical studies.

Preclinical studies were conducted under the context of the PSTC, mostly internally at the FDA or in industry. Conferences were held early in the process with the European Medicines Agency (EMEA) and the Japanese drug regulatory agency. Following the process outlined in the previous section, seven biomarkers were validated and qualified: KIM-1, albumin, total protein, β2-microglobulin, cystatin C, urinary clusterin, and urinary trefoil factor 3.

As a result of the new biomarkers and validation information obtained in these studies, creatinine is no longer sufficient for showing safety at the FDA. The final step in the process occurred in June 2008, when the FDA and EMEA released a statement: "In the first use of a framework allowing submission of a single application to the two agencies, the FDA and the EMEA worked together to allow drug companies to submit the results of seven new tests that evaluate kidney damage during animal studies of new drugs" (FDA, 2008a). The need for better safety biomarkers relating to kidney toxicity and efforts to address this issue are also described in the IOM's recent workshop summary *Accelerating the Development of Biomarkers for Drug Safety* (IOM, 2009a).

Surrogate Endpoints in Nutrition: Foods, Supplements, and Public Health

The following sections describe the types of claims found on food packaging in the United States and how biomarkers play a role in their evidentiary substantiation.

Health Claim Definition

Health claims for foods and dietary supplements are "voluntary statements that characterize the relation between a substance and its ability to reduce the risk of disease or health-related condition" (Schneeman, 2007). Third-party references, written statements, symbols, or vignettes (e.g., brand names including the word heart or heart symbols) that relate a food substance to reduced risk of disease are considered health claims. Implied health claims are statements, symbols, vignettes, and other forms of communication that suggest a relationship between a substance and a disease or health-related condition.[5]

Health claims consist of two parts, a substance (specific food or component of food, including a dietary supplement) and a disease or health-related condition (damage to an organ, part, structure, or system of the body such that it does not function properly or a state of health leading to such dysfunction).[6] In addition, health claims are directed to the general population or population subgroups (e.g., the elderly, women) with the intent to assist the consumer in maintaining healthful dietary practices (CFSAN, 2009a).

As a point of history, prior to the 1990 legislation authorizing health claims, a claim on a food label that referred to a disease condition resulted in the product being classified as a drug and subject to drug regulations. However, emerging science of the 1970s and 1980s had begun to demonstrate a relationship between dietary substances and reduced risk of disease. Taylor and Wilkening (2008) note that "it seemed untenable that only drug products could mention diseases on their labels and even less tenable that food substances with the potential to reduce risk be regulated as drugs." To avoid drug status,[7] health claims cannot assert or imply that they prevent, treat, or mitigate disease, but instead only to reduce the risk of disease.

[5] 21 C.F.R. § 101.14(a)(1) (2008).

[6] 21 C.F.R. § 101.14(a)(2) (2008) and 21 C.F.R. § 101.14(a)(5) (2008).

[7] A drug is defined as an article intended for use in the diagnosis, cure, mitigation, treatment, or prevention of disease. 21 U.S.C. § 321(g)(1)(b).

Legal Basis for Health Claims and Review of Evidence for Health Claims

The Federal Food, Drug, and Cosmetic Act (FDCA) authorizes the Food and Drug Administration to regulate food and dietary supplement labels. In respect to health claims, the FDCA has been amended over time by the 1990 Nutrition Labeling and Education Act (NLEA), the 1994 Dietary Supplement and Health Education Act (DSHEA), and the Food and Drug Administration Modernization Act of 1997. A 1999 court decision (*Pearson* v. *Shalala)* further influenced the FDA's process of evaluating health claims by allowing claims of lesser evidence, accompanied with qualifying language.

The NLEA made nutrition labeling on most foods mandatory and allowed health claims that are based on significant scientific agreement (SSA), or:

> based on the totality of publicly available scientific evidence (including evidence from well-designed studies conducted in a manner which is consistent with generally recognized scientific procedures and principles), that there is significant scientific agreement among experts qualified by scientific training and experience to evaluate such claims, that the claim is supported by such evidence.[8]

DSHEA further amended the FDCA to provide for the use of health claims and nutrient content claims on eligible supplement products, and to provide for the use of structure/function claims. The FDA Modernization Act amended the FDCA to allow health claims based on an authoritative statement of a scientific body of the U.S. government or the National Academy of Sciences. In 1999, the U.S. Court of Appeals found that the SSA standard was overly stringent and violated First Amendment rights by constricting commercial free speech.[9] The court found that claims that did not meet the SSA standard were legal if accompanied by appropriate qualifying language.

In 2009, CFSAN completed a guidance for industry that outlined the agency's current thinking on the process for evaluating scientific evidence for a health claim, the meaning of the SSA standard, and credible scientific evidence to support qualified health claims (CFSAN, 2009a). In the evidence-based review system for the scientific evaluation of health claims, CFSAN has outlined a process to evaluate the strength of the scientific evidence to support a claim about a substance/disease relationship. First, the agency conducts a literature search to identify studies that evaluate the substance/disease relationship, primarily in humans. Studies

[8] 21 C.F.R. § 101.14(c) (1998).

[9] *Pearson* v. *Shalala*, 164 F.3d 650 (D.C. Cir. 1999).

are categorized into intervention studies, observational studies, research synthesis studies, and animal and in vitro studies, and are evaluated and assessed for methodological quality. The agency then sets out to evaluate the totality of scientific evidence about a substance/disease relationship by considering study type, methodological quality rating, number of the various types of studies and sample sizes, relevance of the body of scientific evidence to the U.S. population or target subgroup, replications of findings, and overall consistency of the scientific evidence. Assessing whether the SSA standard is met and specifying the approved claim language are also part of this evidence-based review system.

According to Kathy Ellwood and Paula Trumbo's presentation at the first committee meeting, there is no difference in how the scientific evidence is reviewed for an SSA-level claim or qualified health claim: "Health claims represent a continuum of scientific evidence that extends from very limited or inconclusive evidence to consensus, with evidence supporting SSA health claims lying closer to consensus" (Trumbo and Ellwood, 2009). In the scientific review of evidence for health claims, "Surrogate endpoints of disease risk" considered valid by the FDA's Center for Food Safety and Applied Nutrition include serum LDL cholesterol, total serum cholesterol, and blood pressure for cardiovascular disease; bone mineral density for osteoporosis; adenomatous colon polyps for colon cancer; and elevated blood sugar concentrations and insulin resistance for type 2 diabetes (CFSAN, 2009a). Health claims based on surrogate endpoints include both authorized and qualified claims (Table 2-4). It is important to note that structure/function claims, nutrient content claims, and dietary guidance statements are not based on this scientific evidence review. Because most of these claims do not make reference to disease or health-related conditions, surrogate endpoints are generally not relevant to these types of claims.

Types of Health Claims

Health claims based on significant scientific agreement (authorized health claims) According to Schneeman, the SSA standard "is based on a high level of confidence in the validity of the relation between the substance and the disease or health-related condition" (Schneeman, 2007) and considers the totality of publicly available evidence. When the NLEA was implemented, it required the FDA to consider health claims for 10 specific relationships, of which 8 were approved (Taylor and Wilkening, 2008):

- Calcium and osteoporosis;
- Sodium and hypertension;
- Dietary fat and cancer;

- Dietary saturated fat and cholesterol and CHD;
- Fiber-containing grain products, fruits, and vegetables and cancer;
- Fruits, vegetables, and grain products containing fiber, especially soluble fiber, and CHD;
- Fruits and vegetables and cancer; and
- Folic acid and neural tube defects.

In addition to these initial approved health claims, the NLEA provided a petition process for the consideration of future health claims, involving the petitioner submitting all relevant scientific findings to the FDA. Through this process, an additional seven claims have been approved. Approved health claims that were based on surrogate endpoint data are shown in Table 2-5.

Health claims approved under the SSA standard require specific claim language to be followed. For example, the model health claim language approved for sodium and high blood pressure includes: "Development of hypertension or high blood pressure depends on many factors. [This

TABLE 2-4 Health Claims Based on Surrogate Endpoints

Nutrient	Disease	Surrogate Endpoint	Type of Claim
Phytosterols, soy protein, corn oil, canola oil, and olive oil	Coronary heart disease	LDL and total cholesterol	Phytosterols: Authorized Soy protein: Authorized Corn oil: Qualified Canola oil: Qualified Olive oil: Qualified
Chromium picolinate	Type 2 diabetes	Insulin resistance	Qualified
Calcium and sodium	Hypertension	Systolic and diastolic blood pressure	Calcium: Qualified Sodium: Authorized
Calcium and vitamin D	Osteoporosis	Bone mineral density	Authorized
Calcium	Colorectal cancer	Colorectal polyps	Qualified

NOTE: LDL = low-density lipoprotein.
SOURCE: Trumbo and Ellwood (2009).

TABLE 2-5 Qualified Health Claims Approved by the Food and Drug Administration

Category of Disease	Approved Qualified Health Claims
Cancer	Tomatoes and prostate, ovarian, gastric, and pancreatic cancers
	Calcium and colon/rectal cancer and calcium and colon/rectal polyps
	Green tea and risk of breast, prostate cancer
	Selenium and site-specific cancers
	Antioxidant vitamins C and E and risk of certain cancers
Cardiovascular disease	Folic acid, vitamin B6, vitamin B12 and vascular disease
	Walnuts and coronary heart disease
	Nuts and coronary heart disease
	Omega-3 fatty acids and reduced risk of coronary heart disease
	Corn oil and corn oil-containing products and a reduced risk of heart disease
	Unsaturated fatty acids from canola oil and reduced risk of coronary heart disease
	Monounsaturated fatty acids from olive oil and coronary heart disease
Cognitive function	Phosphatidylserine and cognitive function and dementia
Diabetes	Chromium picolinate and a reduced risk of insulin resistance, type 2 diabetes
Hypertension	Calcium and hypertension, pregnancy-induced hypertension, and preeclampsia
Neural tube defects	Folic acid and neural tube defects

SOURCE: CFSAN (2009b).

product] can be part of a low sodium, low salt diet that might reduce the risk of hypertension or high blood pressure."[10]

Health claims based on authoritative statements The Food and Drug Administration Modernization Act of 1997 specified that the FDA's scientific review process could be circumvented if other scientific bodies of the U.S. government or the National Academy of Sciences[11] had issued authoritative statements about the substance/disease relationship. Authoritative statements from the National Academy of Sciences were

[10] 21 C.F.R. § 101.74(e) (2009).

[11] In legislation, the term National Academy of Sciences refers to the whole of the National Academies.

used to approve three additional health claims—the relationship between whole grains and heart disease, the relationship between certain cancers and potassium, and the relationship between high blood pressure and stroke (Taylor and Wilkening, 2008).

Qualified health claims Litigation over the SSA standard for dietary supplements resulted in an FDA process to approve claims with lesser evidence, given additional qualifying language (qualified health claims). In *Pearson* v. *Shalala*, appellants argued that the high SSA standard impeded First Amendment commercial free speech. According to Schneeman (2007), "courts indicated that the FDA had not presented any data that potentially misleading claim language would not be cured by qualifying language enabling consumers to understand the nature of the evidence supporting a claim." The FDA used a mechanism known as enforcement discretion to allow for the use of qualified health claims (rather than through authorization and publication in the *Federal Register*, as required in the NLEA for SSA health claims) (Taylor and Wilkening, 2008).

As part of a guidance on interim procedures for health claims, FDA proposed a scientific ranking system for health claims, where A-level evidence refers to SSA-level health claims and B-, C-, and D-level evidence refers to the differing levels of evidence for qualified health claims (see Figure 2-5). This ranking system is not used. The FDA approved a B-level qualified health claim for the relationship between walnuts and coronary heart disease. The qualifying language approved was: "supportive but not conclusive research shows that eating 1.5 ounces per day of walnuts, as part of a low saturated fat and low cholesterol diet and not resulting in increased caloric intake, may reduce the risk of coronary heart disease. See nutrition information for fat [and calorie] content" (CFSAN, 2004). The relationship between selenium and cancer was approved as a C-level health claim with the associated qualifying language: "Selenium may reduce the risk of certain cancers. Some scientific evidence suggests that consumption of selenium may reduce the risk of certain forms of cancer. However, [the] FDA has determined that this evidence is limited and not conclusive" (CFSAN, 2003).

An example of qualifying language for a D-level qualified health claim is the relationship between tomatoes/tomato sauce and prostate cancer. The disclaimer language the FDA approved included "very limited and preliminary scientific research suggests that eating one-half to one cup of tomatoes and/or tomato sauce a week may reduce the risk of prostate cancer. [The] FDA concludes that there is little scientific evidence supporting this claim" (CFSAN, 2005). Likewise, the relationship between tomatoes and pancreatic cancer was also approved as a D-level qualified health claim with the associated disclaimer: "one study suggests

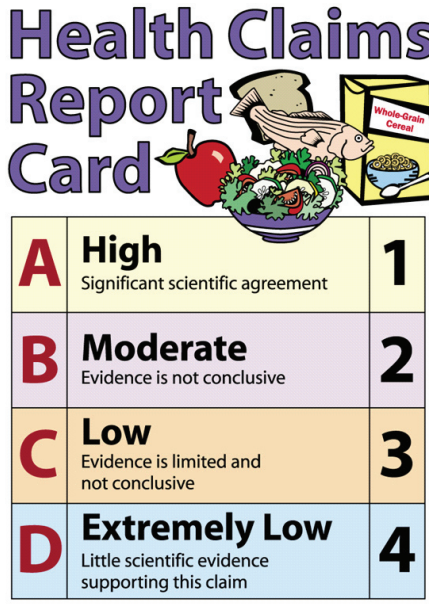

FIGURE 2-5 2003 Food and Drug Administration ranking system for health claims. Claims that met the significant scientific agreement standard were considered A-level claims and were unqualified (requiring no disclaimer). Qualified claims (levels B through D) required disclaimers, such as "evidence is not conclusive."
SOURCES: FDA (2003). See also Mitka (2003).

that consuming tomatoes does not reduce the risk of pancreatic cancer, but one weaker, more limited study suggests that consuming tomatoes may reduce this risk. Based on these studies, [the] FDA concludes that it is highly unlikely that tomatoes reduce the risk of pancreatic cancer" (CFSAN, 2005).

To date, dozens of qualified health claim petitions have been submitted to the FDA. Qualified health claim petitions have been approved for several categories of disease, including cancer, cardiovascular disease, cognitive function, diabetes, hypertension, and neural tube defects (see Table 2-5). On the FDA's website, the denied petitions for qualified health claims are also listed, and include lycopene and cancer, green tea and reduced risk of cardiovascular disease, vitamin E and heart disease, among others (a total of 15 letters of denial have been produced, with one petition—soy protein and cancer—withdrawn) (CFSAN, 2009b).

Other Types of Claims

Nutrient content claims Nutrient content claims expressly or implicitly characterize a level of a nutrient (e.g., "low in fat," "high in vitamin C") in a product (IFT, 2005). Nutrient content claims were established to provide consistent usage throughout the food supply. Prior to the NLEA, nutrient content claims were not standardized, enabling manufacturers to claim "rich in oat bran," "extremely low in saturated fat," with "no assurance that the levels in the food were in fact high or low relative to other similar foods or to an overall diet" (Taylor and Wilkening, 2008).

The FDA currently accepts a number of content claims including free, low, lean, extra lean, high, good source, reduced, less, light, fewer, and more. In addition, the FDA has allowable synonyms for each of the core terms (Taylor and Wilkening, 2008). Nutrient content claims have been authorized for substances that have established Daily Reference Values (DRVs) or Reference Daily Intakes (RDIs), collectively referred to as Daily Values (DVs). For example, a label may claim that the food is "high in," "rich in," or an "excellent source" of a nutrient if the food provides 20 percent or more of the DVs per RACC (Reference Amount Customarily Consumed) (IFT, 2005). Although foods without established DVs cannot have core content claims, manufacturers can make labeling statements, such as "contains *x* mg lycopene per serving," because it does not imply whether the amount of the nutrient is high or low based on DVs, as long as the statement is not misleading (IFT, 2005).

Structure/function claims Claims about the dietary impact of a nutrient on the structure or function of the human body are generally allowed. However, these types of claims cannot suggest that the food or nutrient will cure, mitigate, prevent, or treat disease because that makes it a drug claim. Several structure/function claim examples include "calcium helps build strong bones" or "protein helps build strong muscles." The Institute of Food Technologists note that there is "considerable uncertainty about how far this type of structure/function claim can be 'pushed' before [the] FDA will assert either drug status or health claim status" (IFT, 2005).

Dietary guidance statements Although not considered claims, dietary guidance statements also appear on food labeling. As compared to health claims, dietary guidance statements make reference to either a food substance or a disease, but do not relate these two components in the claim. For example, a dietary guidance statement may say "carrots are good for your health" or "calcium is good for you." Unlike health claims, truthful, non-misleading dietary guidance statements may be used on food labels without premarket review by the FDA (CFSAN, 2008).

BIOMARKERS AND COMMUNICATION STRATEGIES AT THE FDA

Effective use of biomarkers for many purposes depends on the ability of regulators, health-care practitioners, and even advertisers to clearly communicate information about the biomarkers as well as the risks and benefits related to their use. Biomarker use also depends on the ability of the public and others to understand this information. In this and the next section, communication strategies as well as numeracy are discussed, with attention to topics most relevant to public understanding and acceptance of biomarker use.

Research on effective communications in the clinical setting and with respect to prescription and over-the-counter drugs has shown the dramatic effects that good communication strategies can have on patient outcomes. In the clinical setting, studies have pointed to the need for clinicians to receive training on how to communicate with their patients about potential risks of medical treatment (IOM, 2007b; NCI, 2007; Nicholson, 1999). In a review of effective risk communication strategies for cancer genetic counseling, Julian-Reynier and colleagues (2003) emphasized the importance and challenges of providing standardized information about risks of testing to relevant populations as well as individually tailored information based on the patient's immediate concerns. Berry explained many issues of risk communication from a psychology perspective in the book *Risk, Communication and Health Psychology* (Berry, 2004); the understanding and approaches suggested in this book are generally applicable across different health-related settings. The Cochrane Collaboration has reviewed strategies and decision aids for helping patients make decisions about screening tests or health treatments (Edwards et al., 2006; O'Connor et al., 2009). In general, research has found that symbolic representations of probabilistic information, when presented well, are the most effective at enhancing patient–provider communication (Akl et al., 2007; Kim et al., 2009; Lipkus, 2007).

As the primary agency in charge of the safety of foods and drugs, the FDA uses and provides access to a great deal of information on the safety of food, supplements, drugs, biologics, and devices, and on the strength of evidence supporting certain types of health claims on foods and supplements. However, this information can be difficult to access or interpret. Therefore, the main sources of information for clinicians and consumers about the safety, efficacy, and accuracy of product claims that are subject to regulatory review are (1) the labels and package inserts of drugs, biologics, and devices, (2) the drug facts panels found on over-the-counter medication packaging, and (3) the nutrition facts panels and health claims on food packaging.

A recent perspective by Schwartz and Woloshin (2009) in the *New*

England Journal of Medicine highlighted some of the problems with drug labels:

- Drug labels are written by drug companies, and not the FDA. As a result, the FDA may overlook omissions, exaggerations, or inconsistencies in the drug labels.
- For this reason, important information about drug risks may not appear in the final drug label.
- For the same reason, information about the possible benefits of the drugs also may not appear on the drug label.
- A reflection of the reviewers' confidence in the approval decision is rarely reflected in the drug label.

Schwartz and Woloshin noted that the FDA has recognized these problems and has begun to address them. The Risk Communication Advisory Committee was initiated at the FDA in 2008 (see Box 2-4) (FDA, 2008b). A draft guidance not yet finalized was issued in 2006 recommending the use of a prescription drug information highlights panel to "provide immediate access to the information that practitioners most commonly refer to and view as most important" (FDA, 2006b). Inclusion of summaries of the following information was suggested: date of initial U.S. approval, boxed warnings, recent major changes in the label, indications and usage, dosage and administration, dosage forms and strengths, contraindications, warnings and precautions, adverse reactions, drug interactions, and use in special populations.

Effective drug labels have been studied, and the data show that concise, balanced information with symbolic communication aids are useful (Davis et al., 2009; Dowse and Ehlers, 2005; Mansoor and Dowse, 2003; Schwartz et al., 2009). These findings have been discussed at several IOM workshops (IOM, 2007c, 2008), where speakers have suggested that a standardized drug label would improve patient understanding and adherence (IOM, 2008). The challenges of accomplishing this goal were highlighted by Shrank and colleagues (2009) after the conclusion of a study on the ability of a new drug label design to improve patient outcomes in several chronic diseases.

In 2006, the FDA began requiring companies to submit drug label information in an electronic format to enable public access to this information on the FDA website (FDA, 2005). To enhance the usefulness of this information to the public, the committee identified a need to improve the description of the balance of risks and benefits and to expand the product categories included in the database. The website, Drugs@FDA, is not readily found (FDA, 2009). It does not appear on the first 10 pages of

results in a Google search on "FDA electronic drug label," for example.[12] Improvement and expansion of this database and the accessibility of the website would be beneficial.

FURTHER ISSUES WITH USE OF BIOMARKERS

The need for effective communication is important for foods and supplements in addition to drugs, biologics, and devices. A recent report to the FDA Science Board recommended interfacing with universities to improve risk communication (Subcommittee on Science and Technology, 2007). Recommendations 6.1 and 6.2 of the IOM's *The Future of Drug Safety* report focused on ways that the FDA Center for Drug Evaluation and Research could improve risk communication with stakeholders (IOM, 2007a). As a result of these recommendations, the Risk Communication Advisory Committee was created at the FDA. To build on these recommendations, this biomarker evaluation report seeks to extend the intent of these recommendations across regulated product categories and a broader range of stakeholders.

Healthcare providers face a challenging task in conveying health-related information to the public. Professional societies can help healthcare providers obtain skills in how to communicate with their patients about the probabilistic nature of health-related evidence and decisions. Professional societies have an important role to play in helping physicians, consumers, dietitians, other healthcare workers, and individuals in the pharmaceuticals, biologics, medical devices, supplements, and food industries to understand the consequences of innumeracy, evidence gaps, and the insufficiency of evidence to predict all outcomes when evidence is based on surrogate endpoints, other biomarkers, short-term clinical trials, or observational studies alone rather than clinical endpoints.

Numeracy

The need to improve health literacy has been widely recognized. The IOM made recommendations for addressing the issue in a 2004 report in which health literacy was defined as "the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions" (IOM, 2004). That definition had been in use previously by several other groups (HHS, 2000; IOM, 2004; Selden et al., 2000).

One important component of health literacy is numeracy, the ability

[12] Date of the Google search: November 11, 2009. As of March 3, 2010, Drugs@FDA is the second entry on the first page of results.

to understand and interpret the integers, decimals, percentages, and fractions encountered in daily life and to perform related arithmetic (Peters et al., 2007). Its importance actually goes far beyond the ability to understand and make health-related decisions for one's self and family; it is needed for financial transactions, cooking, sewing, building, navigating, and making health-related decisions. Golbeck et al. (2005) define health numeracy as "the degree to which individuals have the capacity to access, process, interpret, communicate, and act on numerical, quantitative, graphical, biostatistical, and probabilistic health information needed to make effective health decisions."

Lower numeracy is associated with less consumer comprehension of drug labels (Davis et al., 2006; Nelson et al., 2008) and food labels (Levy and Fein, 1998; Rothman et al., 2006). Lower numeracy is also associated with poorer health outcomes (Ancker and Kaufman, 2007; Nelson et al., 2008). A great deal of research focuses on strategies for communication between healthcare providers and patients about risks and probabilities (Akl et al., 2007; Apter et al., 2008; Fagerlin et al., 2005; Montori and Rothman, 2005; Peters et al., 2007).

Innumeracy is a problem that goes beyond the general public, however. Researchers have found that numeracy does not necessarily correlate as closely with education as literacy (Jacobson, 2007; Nelson et al., 2008). Nelson et al. (2008) and others recommend the use of short assessments by practitioners so they can better tailor their communication to their patients (Keller and Siergrist, 2009). Furthermore, healthcare practitioners themselves must deal with innumeracy. The adoption and practice of evidence-based medicine depends on physicians' ability to understand and communicate risk and other probabilistic information (Jacobson, 2007; Nusbaum, 2006; Rao, 2008). Innumeracy among other health professionals also needs to be addressed. For example, researchers have examined this issue in nursing (Jukes and Gilchrist, 2006) and psychology (Mulhern and Wylie, 2004).

Numeracy is important to the successful adoption of the biomarker evaluation framework recommended in this report. Understanding biomarker use and the probabilities involved requires comfort with mathematical reasoning. Without adequate numeracy, individuals will have difficulty making decisions under conditions of uncertainty, such as when there are multiple possible outcomes. Without numeracy, regulators will have difficulty explaining to industry the reasoning behind biomarker evaluation, healthcare practitioners will experience difficulty communicating with patients about the probabilities involved with predictions based on biomarkers, and the media will have difficulty in communicating about these topics with the public in general. More work is needed to determine the best ways to communicate probabilistic information and

address innumeracy. The National Research Council has made recommendations on ways to improve numeracy (NRC, 1990, 2005), and the Institute of Medicine has taken several looks at the impact of numeracy on health (IOM, 2001, 2004, 2007b, 2009b). Public support and understanding are important for successful adoption of new policies; informed consumers can help to drive change with respect to careful biomarker evaluation and use.

Cognitive Biases and Impacts of Evidence Gaps

Every day individuals make decisions on the basis of incomplete information on a variety of issues, such as education, safety, diet, health, and more. Although any decision an individual makes may be important in the course of one's life, arguably the decisions related to health are the most likely to affect the length and quality of one's life. For this reason, the stakes are high for these decisions, which are often guided by physicians. But just because the stakes are high does not mean more information is available to use to make an informed decision. Health-related decisions have the same uncertainties as other life decisions. In addition, decisions that policy makers and regulators must make to maximize and protect public health also have these uncertainties. To manage both risks and benefits, all stakeholders—including patients, physicians, and regulatory bodies—need access to reliable information about the uncertainties involved in health decisions.

The goal of access to information can be undermined by the strained resources of government agencies, the overload of information presented to consumers, the profit motivation of companies, and the desire by physicians to reassure their patients. The FDA has a unique relationship with all of these stakeholders and the authority to take actions to protect and promote public health. With better risk communication and access to reliable and complete information about the benefits and risks involved in health decisions, agencies like the FDA will be better able to respond and adjust to the most accurate and current data available for its regulatory decisions.

The committee identified two types of evidence gaps observed when surrogate and other types of biomarkers are used to make decisions about the efficacy of a drug or health benefits of a food. First, they do not explain the entire effect of the food or drug on a person. Second, changes in a biomarker caused by a particular drug, food, or other health intervention do not always predict changes in the clinical outcome of interest. Use of surrogate biomarkers, short-term clinical trials, or observational studies alone cannot adequately predict clinical benefit or harm, and in some cases they do not predict clinical benefit or harm at all. This caution is

even more relevant to decisions based solely on biomarkers whose data do not support use as surrogate endpoints. Without information about an intervention's effect on clinical endpoints, it is impossible to have complete information about the efficacy and safety of the intervention.

Humans tend to oversimplify or ignore evidence gaps in order to make decisions, and are often unaware of evidence gaps. In situations of insufficient or overly complex information, humans use cognitive biases to make decisions; in other words, the types of mistakes people make when making decisions in the absence of complete information are predictable. Tversky and Kahneman explored this area in a famous 1974 paper entitled "Judgment under uncertainty: Heuristics and biases." In this article, Tversky and Kahneman explored the heuristics of representativeness, availability, and anchoring and the biases in judgment that arise from them. Tversky and Kahneman outlined the following heuristics and related cognitive biases in their important 1974 paper:

- The representativeness heuristic (the tendency to make judgments based on how well an element matches to preconceptions of a larger group) leads to the following biases:
  - Insensitivity to prior probability of outcomes (this is also known as neglect of probability bias, or ignoring available probabilistic information when making decisions)
  - Insensitivity to sample size
  - Misconceptions of chance
  - Insentivitiy to predictability (also known as neglect of probability bias, or ignoring available probabilistic information when making decisions)
  - The illusion of validity
  - Misconceptions of regression
- The availability heuristic (making decisions based on the most readily available memories or examples) leads to the following biases:
  - Biases due to the retrievability of instances
  - Biases of imaginability
  - Illusory correlation
- Adjustment and anchoring heuristic (anchoring is the tendency to allow some factor to weigh too heavily in a decision)
  - Insufficient adjustment
  - Biases in the evaluation of conjunctive and disjunctive events
  - Anchoring in the assessment of subjective probability distributions

Each of these heuristics and biases are explained in the referenced paper (Tversky and Kahneman, 1974). An example of insensitivity to probability bias, also known as neglect of probability bias, is when a

person chooses to eat a nutrient or other substance that has been shown in observational studies to be associated with a reduced risk of disease, while ignoring the fact that this research alone does not confirm a substance's causal connection to a reduced risk of disease. Because these biases are well known, some may try to take advantage of them to mislead consumers.

Cognitive biases of healthcare professionals in health-related decision making have been studied in the context of emergent (Pines, 2006), acute (Aberegg et al., 2005; Freshwater-Turner et al., 2007), and chronic healthcare settings (Gruppen et al., 1994; Lutfey and McKinlay, 2009; Redelmeier and Shafir, 1995; Roswarski and Murray, 2006), while cognitive biases of patients have been evaluated in regard to illnesses such as myocardial infarction (Khraim and Carey, 2009) and cancer (Han et al., 2006).

Efforts by professional societies can help physicians, dietitians, and other healthcare practitioners be aware of information gaps and common cognitive biases when helping their patients or clients make decisions about their health care. With this knowledge, strategies can be developed and disseminated. In situations where the public and health professionals need to make decisions in the absence of complete, definitive evidence, decision makers need to be able to access balanced, non-misleading data, or they will be likely to make systematic errors in their thinking.

REFERENCES

Aberegg, S. K., E. F. Haponik, and P. B. Terry. 2005. Omission bias and decision making in pulmonary and critical care medicine. *Chest* 128(3):1497–1505.

Advisory Committee to the Surgeon General. 1964. *Report of the Advisory Committee to the Surgeon General.* Washington, DC: U.S. Department of Health, Education, and Welfare.

Afzal, A. K S. J. Jacobsen, D. W. Mahoney, J. A. Kors, M. M. Redfield, J. C. Burnett, and R. J. Rodeheffer. 2007. Prevalence and prognostic significance of heart failure stages: Application of the American College of Cardiology/American Heart Association heart failure staging criteria in the community. *Circulation* 115(12):1563–1570.

Akl, E. A., N. Maroun, G. Guyatt, A. D. Oxman, P. Alonso-Coello, G. E. Vist, P. J. Devereaux, V. M. Montori, and H. J. Schunemann. 2007. Symbols were superior to numbers for presenting strength of recommendations to health care consumers: A randomized trial. *Journal of Clinical Epidemiology* 60(12):1298–1305.

ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. 2000. Major cardiovascular events in hypertensive patients randomized to doxazosin vs chlorthalidone. *Journal of the American Medical Association* 283(15):1967–1975.

ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. 2002. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs. diuretic: The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). *Journal of the American Medical Association* 288(23):2981–2997.

Alonso, A., G. Molenberghs, H. Geys, M. Buyse, and T. Vangeneugden. 2006. A unifying approach for surrogate marker validation based on Prentice's criteria. *Statistics in Medicine* 25(2):205–221.

Altar, C. A. 2008. The Biomarkers Consortium: On the critical path of drug discovery. *Clinical Pharmacology and Therapeutics* 83(2):361–364.

Altar, C. A., D. Amakye, D. Buonos, J. Bloom, G. Clack, R. Dean, V. Devanarayan, D. Fu, S. Furlong, L. Hinman, C. Girman, C. Lathia, L. Lesko, S. Madani, J. Mayne, J. Meyer, D. Raunig, P. Sager, S. A. Williams, P. Wong, and K. Zerba. 2008. A prototypical process for creating evidentiary standards for biomarkers and diagnostics. *Clinical Pharmacology and Therapeutics* 83(2):368–371.

Ancker, J. S., and D. Kaufman. 2007. Rethinking health numeracy: A multidisciplinary literature review. *Journal of the American Medical Informatics Association* 14(6):713–721.

Apter, A. A., M. K. Paasche-Orlow, J. T. Remillard, I. M. Bennett, E. P. Ben-Joseph, R. M. Batista, H. Hyde, and R. E. Rudd. 2008. Numeracy and communication with patients: They are counting on us. *Journal of General Internal Medicine* 23(12):2117–2124.

Arno, P. S., and K. L. Feiden. 1988. *Against the odds, the story of AIDS drug development, politics and profits*. New York: Harper Collins Publishers.

AVERT. 2009. *History of AIDS: 1987–1992*. http://www.avert.org/aids-history87-92.htm# (accessed October 3, 2009).

Behrman, R. E. 1999. FDA approval of antiretroviral agents: An evolving paradigm. *Drug Information Journal* 33(2):337–341.

Berry, D. C. 2004. *Risk, communication and health psychology*. New York: Open University Press.

Bigger, J. T., J. L. Fleiss, R. Kleiger, J. P. Miller, and L. M. Rolnitzky. 1984. The relationships among ventricular arrhythmias, left ventricular dysfunction, and mortality in the 2 years after myocardial infarction. *Circulation* 69(2):250-258.

Blank, M. 2008. *Clinical Evidentiary Standards for Biomarker Qualification*. Paper presented at 2008 Cardiovascular Biomarkers and Surrogate Endpoints Symposium: Building a Framework for Biomarker Application, Bethesda, MD, September 12.

Boissel, J. P., J. P. Collet, P. Moleur, and M. Haugh. 1992. Surrogate endpoints: A basis for a rational approach. *European Journal of Clinical Pharmacology* 43(3):235–244.

Bonventre, J. V. 2009. *Biomarkers of Kidney Injury: Dawn of a New Era.* Paper read at Institute of Medicine Committee on Qualification of Biomarkers and Surrogate Endpoints in Chronic Disease, Meeting 2, Washington, DC, April 6.

Borer, J. S. 2004. Development of cardiovascular drugs: The U.S. regulatory milieu from the perspective of a participating nonregulator. *Journal of the American College of Cardiology* 44(12):2285–2292.

Brady, T. M., and L. G. Feld. 2009. Pediatric approach to hypertension. *Seminars in Nephrology* 29(4):379–388.

Burzykowski, T., G. Molenberghs, and M. Buyse. 2004. The validation of surrogate end points by using data from randomized clinical trials: A case-study in advanced colorectal cancer. *Journal of the Royal Statistical Society A* 167(Pt 1):103–124.

Buyse, M., and G. Molenberghs. 1998. Criteria for the validation of surrogate endpoints in randomized experiments. *Biometrics* 54(3):1014–1029.

Buyse, M., G. Molenberghs, T. Burzykowski, D. Renard, and H. Geys. 2000. The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics* 1(1):49–67.

CAPS (Cardiac Arrhythmia Pilot Study) Investigators. 1986. The Cardiac Arrhythmia Pilot Study. *American Journal of Cardiology* 57(1):91–95.

CAPS Investigators. 1988. Effects of encainide, flecainide, imipramine and moricizine on ventricular arrrhythmias during the year after acute myocardial infarction: The CAPS. *American Journal of Cardiology* 61(8):501–509.

CAST (Cardiac Arrhythmia Suppression Trial) Investigators. 1989. Preliminary report: Effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. *New England Journal of Medicine* 321(6):406–412.

Carter, B. L. 2002. Blood pressure as a surrogate end point for hypertension. *Annals of Pharmacotherapy* 36(1):87–92.

CBER (Center for Biologics Evaluation and Research). 2007. *Guidance for industry: Clinical data needed to support the licensure of seasonal inactivated influenza vaccines*. http://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidances/Vaccines/ucm074794.htm (accessed September 24, 2009).

CDRH (Center for Devices and Radiological Health). 2006. *Guidance for industry and FDA staff: Postmarket surveillance under section 522 of the Federal Food, Drug, and Cosmetic Act*. http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm072517.htm (accessed September 24, 2009).

CDRH. 2008. *Guidance for industry and FDA staff: Clinical study designs for catheter ablation devices for treatment of atrial flutter*. http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm070919.htm (accessed November 10, 2009).

CFSAN (Center for Food Safety and Applied Nutrition). 2003. *Selenium and certain cancers (Qualified health claim: Final decision letter)*. http://www.cfsan.fda.gov/~dms/ds-ltr35.html (accessed March 17, 2009).

CFSAN. 2004. *Qualified health claims: Letter of enforcement discretion—walnuts and coronary heart disease* (accessed March 17, 2009).

CFSAN. 2005. *Qualified health claims: Letter regarding "Tomatoes and prostate, ovarian, gastric and pancreatic cancers (American Longevity Petition)."* http://www.cfsan.fda.gov/~dms/qhclyco.html (accessed March 17, 2009).

CFSAN. 2008. *A food labeling guide*. http://www.cfsan.fda.gov/~dms/2lg-8.html#health (accessed March 26, 2009).

CFSAN. 2009a. *Guidance for industry: Evidence-based review system for the scientific evaluation of health claims*. http://www.cfsan.fda.gov/~dms/hclmgui6.html (accessed February 25, 2009).

CFSAN. 2009b. *Qualified health claims*. http://www.foodsafety.gov/~dms/lab-qhc.html (accessed March 17, 2009).

Clarke, R., S. Lewington, P. Sherliker, and J. Armitage. 2007. Effects of B-vitamins on plasma homocysteine concentrations and on risk of cardiovascular disease and dementia. *Current Opinion in Clinical Nutrition and Metabolic Care* 10(1):32–39.

Cochrane, A. L. 1971. *Effectiveness and efficiency: Random reflections on health services.* London, England: Nuffield Provincial Hospitals Trust.

Colatsky, T. J. 2009. Reassessing the validity of surrogate markers of drug efficacy in the treatment of coronary artery disease. *Current Opinion in Investigational Drugs* 10(3):239–244.

Colburn, W. A. 1997. Selecting and validating biologic markers for drug development. *Journal of Clinical Pharmacology* 37(5):355–362.

Colburn, W. A. 2000. Optimizing the use of biomarkers, surrogate endpoints, and clinical endpoints for more efficient drug development. *Journal of Clinical Pharmacology* 40(12 Pt 2):1419–1427.

Cotton, P. 1991. HIV surrogate markers weighed. *Journal of the American Medical Association* 265(11):1357, 1361–1362.

Daniels, M. J., and M. D. Hughes. 1997. Meta-analysis for the evaluation of potential surrogate markers. *Statistics in Medicine* 16(17):1965–1982.

Davis, T. C., M. S. Wolf, P. F. Bass, J. A. Thompson, H. H. Tilson, M. Neuberger, and R. M. Parker. 2006. Literacy and misunderstanding prescription drug labels. *Annals of Internal Medicine* 145(12):887–894.

Davis, T. C., A. D. Federman, P. F. Bass, R. H. Jackson, M. Middlebrooks, R. M. Parker, and M. S. Wolf. 2009. Improving patient understanding of prescription drug label instructions. *Journal of General Internal Medicine* 24(1):57–62.

De Gruttola, V., T. Fleming, D. Y. Lin, and R. Coombs. 1997. Perspective: Validating surrogate markers—Are we being naive? *Journal of Infectious Diseases* 175(2):237–246.

De Gruttola, V., C. Flexner, J. Schapiro, M. Hughes, M. van der Laan, and D. R. Kuritzkes. 2006. Drug development strategies for salvage therapy: Conflicts and solutions. *AIDS Research and Human Retroviruses* 22(11):1106–1109.

Deeks, S. G., and J. N. Martin. 2007. Partial treatment interruptions. *Current Opinion in HIV and AIDS* 2(1):46–55.

DeMets, D. L., and R. M. Califf. 2002. Lessons learned from recent cardiovascular clinical trials: Part I. *Circulation* 106(6):746–751.

Desai, M., N. Stockbridge, and R. Temple. 2006. Blood pressure as an example of a biomarker that functions as a surrogate. *AAPS Journal* 8(1):E146.

Dowse, R., and M. Ehlers. 2005. Medicine labels incorporating pictograms: Do they influence understanding and adherence? *Patient Education and Counseling* 58(1):63–70.

Echt, D. S., P. R. Liebson, L. B. Mitchell, R. W. Peters, D. Obias-Manno, A. H. Barker, D. Arensberg, A. Baker, L. Friedman, H. L. Greene, M. L. Huther, D. W. Richardson, and the CAST Investigators. 1991. Mortality and morbidity in patients receiving encainide, flecainide, or placebo: The Cardiac Arrhythmia Suppression Trial. *New England Journal of Medicine* 324(12):781-788.

Edwards, A. G. K., R. Evans, J. Dundon, S. Haigh, K. Hood, and G. J. Elwyn. 2006. Personalised risk communication for informed decision making about taking screening tests. *Cochrane Database of Systematic Reviews* 2006(4):Art. No. CD001865.

Ellenberg, S. S., and J. M. Hamilton. 1989. Surrogate endpoints in clinical trials: Cancer. *Statistics in Medicine* 8(4):405–413.

Emanuel, E. J., and F. G. Miller. 2001. The ethics of placebo-controlled trials—A middle ground. *New England Journal of Medicine* 345(12):915–919.

Ezzati, M., A. D. Lopez, A. Rodgers, S. V. Hoorn, C. J. L. Murray, and C. R. A. C. Group. 2002. Selected major risk factors and global and regional burden of disease. *The Lancet* 360(9343):1347–1360.

Fagerlin, A., C. Wang, and P. A. Ubel. 2005. Reducing the influence of anecdotal reasoning on people's health care decisions: Is a picture worth a thousand statistics? *Medical Decision Making* 25(4):398–405.

FDA (Food and Drug Administration). 1988. *Making drugs available for life-threatening diseases*. http://www.aegis.com/news/FDA/1988/Fd881001.html (accessed October 3, 2009).

FDA. 1999. *Significant scientific agreement in the review of health claims for conventional foods and dietary supplements*. http://www.fda.gov/Food/GuidanceComplianceRegulatoryInformation/GuidanceDocuments/FoodLabelingNutrition/ucm059132.htm (accessed October 3, 2009).

FDA. 2002. *Guidance for industry: Antiretroviral drugs using plasma HIV RNA measurements—Clinical considerations for accelerated and traditional approval*. http://www.fda.gov/cder/guidance/3647fnl.pdf (accessed December 29, 2008).

FDA. 2003. *FDA to encourage science-based labeling and competition for healthier dietary choices*. http://www.fda.gov/bbs/topics/NEWS/2003/NEW00923.html (accessed March 17, 2009).

FDA. 2004. *Guidance for industry: Available therapy*. http://www.fda.gov/RegulatoryInformation/Guidances/ucm126586.htm (accessed September 24, 2009).

FDA. 2005a. *FDA announces the use of new electronic drug labels to help better inform the public and improve patient safety.* http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/2005/ucm108509.htm (accessed November 22, 2009).

FDA. 2006a. *Guidance for industry: Clinical studies section of labeling for human prescription drug and biological products—Content and format*. http://www.fda.gov/RegulatoryInformation/Guidances/ucm127509.htm (accessed September 24, 2009).

FDA. 2006b. *Draft guidance for industry: Labeling for human prescription drug and biological products—Implementing the new content and format requirements*. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm075082.pdf (accessed November 22, 2009).

FDA. 2007. *Guidance for industry: Clinical trial endpoints for the approval of cancer drugs and biologics*. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM071590.pdf (accessed September 24, 2009).

FDA. 2008a. *FDA, European Medicines Agency to consider additional test results when assessing new drug safety*. http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/2008/ucm116911.htm (accessed November 23, 2009).

FDA. 2008b. *Risk Communication Advisory Committee*. http://www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/RiskCommunicationAdvisoryCommittee/default.htm (accessed November 21, 2009).

FDA. 2009. *Drugs@FDA*. http://www.accessdata.fda.gov/Scripts/cder/DrugsatFDA/ (accessed November 11, 2009).

FDA. 2010. *Risk Communication Advisory Committee*. http://www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/RiskCommunicationAdvisoryCommittee/default.htm (accessed March 14, 2010).

Fischl, M. A., D. D. Richman, M. H. Grieco, M. S. Gottlieb, P. A. Volberding, O. L. Laskin, J. M. Leedom, J. E. Groopman, D. Mildvan, R. T. Schooley et al. 1987. The efficacy of azidothymidine (AZT) in the treatment of patients with AIDS and AIDS-related complex. A double-blind, placebo-controlled trial. *New England Journal of Medicine* 317(4):185–191.

Fleming, T. R. 2005. Surrogate endpoints and FDA's accelerated approval process. *Health Affairs* 24(1):67–78.

Fleming, T. R., and D. L. DeMets. 1996. Surrogate end points in clinical trials: Are we being misled? *Annals of Internal Medicine* 125(7):605–613.

Frangakis, C. E., and D. B. Rubin. 2002. Principal stratification in causal inference. *Biometrics* 58(1):21–29.

Freshwater-Turner, D. A., R. J. Boots, R. N. Bowman, H. G. Healy, and A. C. Klestov. 2007. Difficult decisions in the intensive care unit: An illustrative case. *Anaesthesia and Intensive Care* 35(5):748–759.

Gobburu, J. V. S. 2009. Biomarkers in clinical drug development. *Clinical Pharmacology and Therapeutics* 86(1):26–27.

Golbeck, A. L., C. R. Ahlers-Schmidt, A. M. Paschal, and S. E. Dismuke. 2005. A definition and operational framework for health numeracy. *American Journal of Preventive Medicine* 29(4):375–376.

Goodsaid, F. 2008a. *Impact of the Biomarker Qualification Project on Drug Development*. Paper presented at 2008 Cardiovascular Biomarkers and Surrogate Endpoints Symposium: Building a Framework for Biomarker Application, Bethesda, MD, September 12.

Goodsaid, F. 2008b. *Markers of Renal Toxicity*. Paper presented at 2008 Cardiovascular Biomarkers and Surrogate Endpoints Symposium: Building a Framework for Biomarker Application, Bethesda, MD, September 11.

Goodsaid, F., and F. Frueh. 2006. Process map proposal for the validation of genomic biomarkers. *Pharmacogenomics* 7(5):773–782.

Goodsaid, F., and F. Frueh. 2007a. Biomarker qualification pilot process at the U.S. Food and Drug Administration. *AAPS Journal* 9(1):E105–E108.

Goodsaid, F., and F. Frueh. 2007b. Questions and answers about the pilot process for biomarker qualification at the FDA. *Drug Discovery Today: Technologies* 4(1):9–11.

Goodsaid, F., F. W. Frueh, and W. Mattes. 2008. Strategic paths for biomarker qualification. *Toxicology* 245:219–223.

Gruppen, L. D., J. Margolin, K. Wisdom, and C. M. Grum. 1994. Outcome bias and cognitive dissonance in evaluating treatment decisions. *Academic Medicine* 69(10 Suppl): S57–S59.

Hammer, S. M., K. E. Squires, M. D. Hughes, J. M. Grimes, L. M. Demeter, J. S. Currier, J. J. Eron Jr., J. E. Feinberg, H. H. Balfour Jr., L. R. Deyton, J. A. Chodakewitz, and M. A. Fischl. 1997. A controlled trial of two nucleoside analogues plus indinavir in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less. AIDS Clinical Trials Group 320 Study Team. *New England Journal of Medicine* 337(11):725–733.

Han, P. K., R. P. Moser, and W. M. Klein. 2006. Perceived ambiguity about cancer prevention recommendations: Relationship to perceptions of cancer preventability, risk, and worry. *Journal of Health Communication* 11(Suppl 1):51–69.

HHS (Department of Health and Human Services). 2000. *Healthy People 2010: Understanding and improving health*. Washington, DC: HHS.

Higgins, J., and S. Green. 2008. *Cochrane handbook for systematic reviews of interventions: Version 5.0.1*. http://www.cochrane.org/resources/handbook/index.htm (accessed April 2, 2009).

Hill, A. B. 1965. The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine* 58:295–300.

Hillis, A., and D. Seigel. 1989. Surrogate endpoints in clinical trials: Ophthalmologic disorders. *Statistics in Medicine* 8(4):427–430.

HIV Surrogate Marker Collaborative Group. 2000. Human immunodeficiency virus type 1 RNA level and CD4 count as prognostic markers and surrogate end points: A meta-analysis. *AIDS Research and Human Retroviruses* 16(12):1123–1133.

Holden, C. 1993. FDA okays surrogate markers. *Science* 259(5091):32–33.

Hughes, M. D. 2002. Evaluating surrogate endpoints. *Controlled Clinical Trials* 23(6): 703–707.

Hughes, M. D. 2005. The evaluation of surrogate endpoints in practice: Experience in HIV. In *The evaluation of surrogate endpoints*, edited by T. Burzykowski, G. Molenberghs, and M. Buyse. New York: Springer. Pp. 295–321.

Hughes, M. D., V. De Gruttola, and S. L. Welles. 1995. Evaluating surrogate markers. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* 10(Suppl 2):S1–S8.

Hughes, M. D., M. J. Daniels, M. A. Fischl, S. Kim, and R. T. Schooley. 1998. CD4 cell count as a surrogate endpoint in HIV clinical trials: A meta-analysis of studies of the AIDS Clinical Trials Group. *AIDS* 12(14):1823–1832.

IFT (Institute of Food Technologists). 2005. *Functional foods: Opportunities and challenges*. Chicago, IL: IFT.

IOM (Institute of Medicine). 1990. *Clinical practice guidelines: Directions for a new program*. Washington, DC: National Academy Press.

IOM. 2001. *Health and behavior: The interplay of biological, behavioral, and societal influences*. Washington, DC: National Academy Press.

IOM. 2004. *Health literacy: A prescription to end confusion*. Washington, DC: The National Academies Press.

IOM. 2005. *Saving women's lives: Strategies for improving breast cancer detection and diagnosis*. Washington, DC: The National Academies Press.

IOM. 2007a. *The future of drug safety: Promoting and protecting the health of the public*. Washington, DC: The National Academies Press.

IOM. 2007b. *Understanding the benefits and risks of pharmaceuticals: Workshop Summary*. Washington, DC: The National Academies Press.

IOM. 2008. *Standardizing medication labels: Confusing patients less: Workshop summary*. Washington, DC: National Academies Press.

IOM. 2009a. *Accelerating the development of biomarkers for drug safety*. Washington, DC: The National Academies Press.

IOM. 2009b. *Health literacy, eHealth, and communication: Putting the consumer first: Workshop Summary*. Washington, DC: The National Academies Press.

IOM. 2009c. *Initial national priorities for comparative effectiveness research*. Washington, DC: The National Academies Press.

Israili, Z. H., R. Hernandez-Hernandez, and M. Valasco. 2007. The future of antihypertensive treatment. *American Journal of Therapeutics* 14(2):121–134.

Jacobson, R. M. 2007. Teaching numeracy to physicians-in-training: Quantitative analysis for evidence-based medicine. *Minnesota Medicine* 90(11):37–38, 46.

Jiang, H., S. G. Deeks, D. R. Kuritzkes, M. Lallemant, D. Katzenstein, M. Albrecht, and V. De Gruttola. 2003. Assessing resistance costs of antiretroviral therapies via measures of future drug options. *Journal of Infectious Diseases* 188(7):1001–1008.

Joffe, M. M., and T. Greene. 2009. Related causal frameworks for surrogate outcomes. *Biometrics* 65(2):530–538.

Jukes, L., and M. Gilchrist. 2006. Concerns about numeracy skills of nursing students. *Nurse Education in Practice* 6(4):192–198.

Julian-Reynier, C., M. Welkenhuysen, L. Hagoel, M. Decruyenaere, P. Hopwood, and on behalf of the CRISCOM Working Group. 2003. Risk communication strategies: State of the art and effectiveness in the context of cancer genetic services. *European Journal of Human Genetics* 11(10):725–736.

Keller, C., and M. Siergrist. 2009. Effect of risk communication formats on risk perception depending on numeracy. *Medical Decision Making* 29(4):483–490.

Khraim, F. M., and M. G. Carey. 2009. Predictors of pre-hospital delay among patients with acute myocardial infarction. *Patient Education and Counseling* 75(2):155–161.

Kim, H., C. Nakamura, and Q. Zeng-Treitler. 2009. Assessment of pictographs developed through a participatory design process using an online survey tool. *Journal of Medical Internet Research* 11(1):e5.

Krumholz, H. M., and T. H. Lee. 2008. Redefining quality—Implications of recent clinical trials. *New England Journal of Medicine* 358(24):2537–2539.

Lagakos, S. W., and D. F. Hoth. 1992. Surrogate markers in AIDS: Where are we? Where are we going? *Annals of Internal Medicine* 116(7):599–601.

Lassere, M. N. 2008. The biomarker-surrogacy evaluation schema: A review of the biomarker-surrogate literature and a proposal for a criterion-based, quantitative, multidimensional hierarchical levels of evidence schema for evaluating the status of biomarkers as surrogate endpoints. *Statistical Methods in Medical Research* 17(3):303–340.

Lathia, C. D., D. Amakye, W. Dai, C. Girman, S. Madani, J. Mayne, P. MacCarthy, P. Pertel, L. Seman, A. Stoch, P. Tarantino, C. Webster, S. Williams, and J. A. Wagner. 2009. The value, qualification, and regulatory use of surrogate end points in drug development. *Clinical Pharmacology and Therapeutics* 86(1):32–43.

Law, M. R., and J. K. Morris. 2009. Use of blood pressure lowering drugs in the prevention of cardiovascular disease: Meta-analysis of 147 randomized trials in the context of expectations from prospective epidemiological studies. *British Medicine Journal* 338:b1665.

Lesko, L. J., and A. J. Atkinson Jr. 2001. Use of biomarkers and surrogate endpoints in drug development and regulatory decision making: Criteria, validation, strategies. *Annual Review of Pharmacology and Toxicology* 41:347–366.

Levy, A. S., and S. B. Fein. 1998. Consumers' ability to perform tasks using nutrition labels. *Journal of Nutrition Education* 30(4):210–217.

Lipkus, I. M. 2007. Numeric, verbal, and visual formats of conveying health risks: Suggested best practices and future recommendations. *Medical Decision Making* 27(5):696–713.

Lutfey, K. E., and J. B. McKinlay. 2009. What happens along the diagnostic pathway to CHD treatment? Qualitative results concerning cognitive processes. *Sociology of Health and Illness* 31(7):1077–1092.

MacMahon, S., R. Peto, J. Cutler, R. Collins, P. Sorlie, J. Neaton, R. Abbott, J. Godwin, A. Dyer, and J. Stamler. 1990. Blood pressure, stroke, and coronary heart disease. Part 1, prolonged differences in blood pressure: Prospective observational studies corrected for the regression dilution bias. *Lancet* 335(8692):765–774.

Manns, B., W. F. Owen, W. C. Winkelmayer, P. J. Deveraux, and M. Tonelli. 2006. Surrogate markers in clinical studies: Problems solved or created? *American Journal of Kidney Diseases* 48(1):159–166.

Mansoor, L. E., and R. Dowse. 2003. Effect of pictograms on readability of patient information materials. *Annals of Pharmacotherapy* 37(7–8):1003–1009.

Miksad, R. 2009. *Decision science for qualification of biomarker surrogate endpoints*. Paper read at Institute of Medicine Committee on Qualification of Biomarkers and Surrogate Endpoints in Chronic Disease, Meeting 2, Washington, DC, April 6.

Mitka, M. 2003. Food fight over product label claims: Critics say proposed changes will confuse consumers. *Journal of the American Medical Association* 290(7):871–875.

Montori, V. M., and R. L. Rothman. 2005. Weakness in numbers: The challenge of numeracy in health care. *Journal of General Internal Medicine* 20(11):1071–1072.

Mukharji, J., R. E. Rude, W. K. Poole, N. Gustafson, L. J. Thomas Jr., H. W. Strauss, A. S. Jaffe, J. E. Muller, R. Roberts, D. S. Raabe, C. H. Croft, E. Passamani, E. Braunwald, J. T. Willerson, and the MILIS Study Group. 1984. Risk factors for sudden death after acute myocardial infarction: Two-year follow-up. *American Journal of Cardiology* 54(1):31–36.

Mulhern, G., and J. Wylie. 2004. Changing levels of numeracy and other core mathematical skills among psychology undergraduates between 1992 and 2002. *British Journal of Psychology* 95(Pt 3):355–370.

Nambi, V., and C. M. Ballantyne. 2007. Role of biomarkers in developing new therapies for vascular disease. *World Journal of Surgery* 31(4):676–681.

NCI (National Cancer Institute). 2007. *Patient-centered communication in cancer care*. Washington, DC: NCI.

Nelson, W., V. F. Reyna, A. Fagerlin, I. Lipkus, and E. Peters. 2008. Clinical implications of numeracy: Theory and practice. *Annals of Behavioral Medicine* 35(3):261–274.

NHLBI Working Group on Future Directions in Hypertension Treatment Trials. 2005. Major clinical trials of hypertension: What should be done next? *Hypertension* 46(1): 1–6.

Nicholson, P. J. 1999. Communicating health risk. *Occupational Medicine* 49(4):253–256.

NRC (National Research Council). 1990. *On the shoulders of giants: New approaches to numeracy*. Washington, DC: National Academy Press.

NRC. 2005. *Measuring literacy: Performance levels for adults*. Washington, DC: The National Academies Press.

Nusbaum, N. J. 2006. Mathematics preparation for medical school: Do all premedical students need calculus? *Teaching and Learning in Medicine* 18(2):165–168.

O'Connor, A. M., C. L. Bennett, D. Stacey, M. Barry, N. F. Col, K. B. Eden, V. A. Entwistle, V. Fiset, M. Holmes-Rovner, S. Khangura, H. Llewellyn-Thomas, and D. Rovner. 2009. Decision aids for people facing health treatment or screening decisions. *Cochrane Database of Systematic Reviews* 2009(3): Art. No. CD001431.

Omenn, G. S., G. E. Goodman, M. D. Thornquist, J. Balmes, M. R. Cullen, A. Glass, J. P. Keogh, F. L. Meyskens Jr., B. Valanis, J. H. Williams Jr., S. Barnhart, and S. Hammar. 1996. Effects of a combination of beta-carotene and vitamin A on lung cancer and cardiovascular disease. *New England Journal of Medicine* 334:1150–1155.

Perry, H. M., A. I. Goldman, M. A. Lavin, H. W. Schnaper, A. E. Fitz, E. D. Frohlich, B. Steele, and H. G. Richman. 1978. Evaluation of drug treatment in mild hypertension: VA-NHLBI feasibility trial. Plan and preliminary results of a two-year feasibility trial for a multicenter intervention study to evaluate the benefits versus the disadvantages of treating mild hypertension. Prepared for the Veterans Administration–National Heart, Lung, and Blood Institute Study Group for Evaluating Treatment in Mild Hypertension. *Annals of the New York Academy of Sciences* 304:267–287.

Peters, E., J. Hibbard, P. Slovic, and N. Dieckmann. 2007. Numeracy skill and the communication, comprehension, and use of risk–benefit information. *Health Affairs* 26(3):741–748.

Peto, R., R. Doll, J. D. Buckley, and M. B. Sporn. 1981. Can dietary beta-carotene materially reduce human cancer rates? *Nature* 290:201–208.

Pines, J. M. 2006. Profiles in patient safety: Confirmation bias in emergency medicine. *Academic Emergency Medicine* 13(1):90–94.

Prentice, R. L. 1989. Surrogate endpoints in clinical trials: Definition and operational criteria. *Statistics in Medicine* 8(4):431–440.

Psaty, B. M., D. S. Siscovick, N. S. Weiss, T. D. Koepsell, F. R. Rosendaal, D. Lin, S. R. Heckbert, E. H. Wagner, and C. D. Furberg. 1996. Hypertension and outcomes research: From clinical trials to clinical epidemiology. *American Journal of Hypertension* 9(2):178–183.

Rao, G. 2008. Physician numeracy: Essential skills for practicing evidence-based medicine. *Family Medicine* 40(5):354–358.

Redelmeier, D. A., and E. Shafir. 1995. Medical decision making in situations that offer multiple alternatives. *Journal of the American Medical Association* 273(4):302–305.

Robins, J. M., and S. Greenland. 1992. Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 3(2):143–155.

Robins, J. M., and S. Greenland. 1994. Adjusting for differential rates of prophylaxis therapy for PCP in high- versus low-dose AZT treatment arms in an AIDS randomized trial. *Journal of the American Statistical Association* 89(427):737–749.

Roswarski, T. E., and M. D. Murray. 2006. Supervision of students may protect academic physicians from cognitive bias: A study of decision making and multiple treatment alternatives in medicine. *Medical Decision Making* 26(2):154–161.

Rothman, R. L., R. Housam, H. Weiss, D. Davis, R. Gregory, T. Gebretsadik, A. Shintani, and T. A. Elasy. 2006. Patient understanding of food labels—The role of literacy and numeracy. *American Journal of Preventive Medicine* 31(5):391–398.

Ruberman, W., E. Weinblatt, J. D. Goldberg, C. W. Frank, and S. Shapiro. 1977. Ventricular premature beats and mortality after myocardial infarction. *New England Journal of Medicine* 297(14):750–757.

Ruskin, J. N. 1989. The Cardiac Arrhythmia Suppression Trial (CAST). *New England Journal of Medicine* 321(6):386–388.

Schatzkin, A., and M. Gail. 2002. The promise and peril of surrogate endpoints in cancer research. *Nature Reviews Cancer* 2(1):19–27.

Schneeman, B. 2007. FDA's review of scientific evidence for health claims. *Journal of Nutrition* 137(2):493–494.

Schwartz, L. M., S. Woloshin, and H. G. Welch. 2009. Using a drug facts box to communicate drug benefits and harms: Two randomized trials. *Annals of Internal Medicine* 150(8):516–527.

Selden, C. R., M. Zorn, S. C. Ratzan, and R. M. Parker. 2000. *National Library of Medicine current bibliographies in medicine: Health literacy*. NLM Pub. No. CBM 2000-1. Bethesda, MD: National Institutes of Health.

Shi, Q., and D. J. Sargent. 2009. Meta-analysis for the evaluation of surrogate endpoints in cancer clinical trials. *International Journal of Clinical Oncology* 14(2):102–111.

Shrank, W. H., A. Patrick, P. P. Gleason, C. Canning, C. Walters, A. H. Heaton, S. Jan, M. A. Brookhart, S. Schneeweiss, D. H. Solomon, M. S. Wolf, J. Avorn, and N. K. Choudhry. 2009. An evaluation of the relationship between the implementation of a newly designed prescription drug label at Target pharmacies and health outcomes. *Medical Care* 47(9):1031–1035.

Subcommittee on Science and Technology. 2007. *FDA science and mission at risk.* Washington, DC: Food and Drug Administration.

Taylor, C. L., and V. L. Wilkening. 2008. How the nutrition food label was developed, part 2: The purpose and promise of nutrition claims. *Journal of the American Dietetic Association* 108(4):618–623.

Temple, R. 1999. Are surrogate markers adequate to assess cardiovascular disease drugs? *Journal of the American Medical Association* 282(8):790–795.

Temple, R. J. 2009. *Qualification of Biomarkers as Surrogate Endpoints of Chronic Disease Risk.* Paper read at Institute of Medicine Committee on Qualification of Biomarkers and Surrogate Endpoints in Chronic Disease: Meeting 2, Washington, DC, April 6.

Trumbo, P., and K. Ellwood. 2009. *Developing a Framework for Biomarker Qualification for Chronic Disease*. Paper presented at Institute of Medicine Committee on Qualification of Biomarkers as Surrogate Endpoints for Chronic Disease Risk, Washington, DC, January 12.

Tversky, A. And D. Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185(4157):1124–1131.

Volberding, P. A., S. W. Lagakos, J. M. Grimes, D. S. Stein, H. H. Balfour Jr., R. C. Reichman, J. A. Bartlett, M. S. Hirsch, J. P. Phair, R. T. Mitsuyasu et al. 1994. The duration of zidovudine benefit in persons with asymptomatic HIV infection. Prolonged evaluation of protocol 019 of the AIDS Clinical Trials Group. *Journal of the American Medical Association* 272(6):437–442.

Wagner, J. A. 2002. Overview of biomarkers and surrogate endpoints in drug development. *Disease Markers* 18(2):41–46.

Wagner, J. A. 2008. Strategic approach to fit-for-purpose biomarkers in drug development. *Annual Review of Pharmacology and Toxicology* 48:631–651.

Wagner, J. A., S. A. Williams, and C. J. Webster. 2007. Biomarkers and surrogate end points for fit-for-purpose development and regulatory evaluation of new drugs. *Clinical Pharmacology and Therapeutics* 81(1):104–107.

Wang, T. J., and R. S. Vasan. 2005. Epidemiology of uncontrolled hypertension in the United States. *Circulation* 112(11):1651–1662.

Williams, B. 2005. Recent hypertension trials. *Journal of the American College of Cardiology* 45(6):813–827.

Williams, S. A., D. E. Slavin, J. A. Wagner, and C. J. Webster. 2006. A cost-effectiveness approach to the qualification and acceptance of biomarkers. *Nature Reviews Drug Discovery* 5(11):897–902.

Wittes, J., E. Lakatos, and J. Probstfield. 1989. Surrogate endpoints in clinical trials: Cardiovascular diseases. *Statistics in Medicine* 8(4):415–425.

Wolff, T., and T. Miller. 2007. Evidence for the reaffirmation of the U.S. Preventive Services Task Force recommendation on screening for high blood pressure. *Annals of Internal Medicine* 147(11):787–791.

3

The Biomarker Evaluation Process

The previous chapter's detailed exploration of biomarker evaluation efforts indicates a need for a unified, transparent process for the evaluation and adoption of biomarkers. Although the principal purpose for evaluation is to ensure that a biomarker is scientifically and clinically meaningful for specified purposes (Palou et al., 2009; Wagner, 2008; Wagner et al., 2007; Williams et al., 2006), evaluation also allows for informed decisions about which biomarkers to pursue and data to gather. This chapter begins to present the committee's recommendations on the best ways to proceed (see Box 3-1 for the recommendations discussed in this chapter).

The committee's biomarker evaluation framework was informed by the previously developed qualification frameworks discussed in Chapter 2; the committee determined there are three necessary components to biomarker evaluation: (1) analytical validation of relevant biomarker tests; (2) qualification, a description of the evidence relating to the biomarker in question—as measured using validated tests—to the intervention and disease outcome; and (3) utilization, the applicability of results from the analytical validation and the description of the evidence to the proposed use of the biomarker given the evidence assessment and proposed purpose and context of use. Thus, the committee's framework has three distinct yet interrelated steps; they are not necessarily separated in time (i.e., some of the steps may occur concurrently) and conclusions in one step may require revisions or additional work in other steps (see Figure 3-1). Previous evaluation frameworks have not explicitly incorporated a process for reevaluating the three steps of the biomarker assessments based on new

BOX 3-1
Recommendations 1–4

Recommendation 1:

The biomarker evaluation process should consist of the following three steps:

1a. Analytical validation: analyses of available evidence on the analytical performance of an assay;

1b. Qualification: assessment of available evidence on associations between the biomarker and disease states, including data showing effects of interventions on both the biomarker and clinical outcomes; and

1c. Utilization: contextual analysis based on the specific use proposed and the applicability of available evidence to this use. This includes a determination of whether the validation and qualification conducted provide sufficient support for the use proposed.

Recommendation 2:

2a. For biomarkers with regulatory impact, the Food and Drug Administration (FDA) should convene expert panels to evaluate biomarkers and biomarker tests.

2b. Initial evaluation of analytical validation and qualification should be conducted separately from a particular context of use.

2c. The expert panels should reevaluate analytical validation, qualification, and utilization on a continual and a case-by-case basis.

Recommendation 3:

The FDA should use the same degree of scientific rigor for evaluation of biomarkers across regulatory areas, whether they are proposed for use in the arenas of drugs, medical devices, biologics, or foods and dietary supplements. Congress may need to strengthen FDA authority to accomplish this goal.

Recommendation 4:

The FDA should take into account a nutrient's or food's source as well as any modifying effects of the food or supplement that serves as the delivery vehicle and the dietary patterns associated with consumption of the nutrient or food when reviewing health-related label claims and the safety of food and supplements. Congress may need to strengthen FDA authority to accomplish this goal.

data; the committee's framework explicitly includes such a process, while allowing for timely, reliable, and effective decision making.

The evaluation framework is intended to be applicable across a wide range of biomarker uses, from exploratory uses for which less evidence is required to surrogate endpoint uses for which compelling evidence is required. The framework is meant for, but not limited to, use in research, clinical, product, and claim development in food, drug, and device industries as well as public health settings, and it is intended to function for

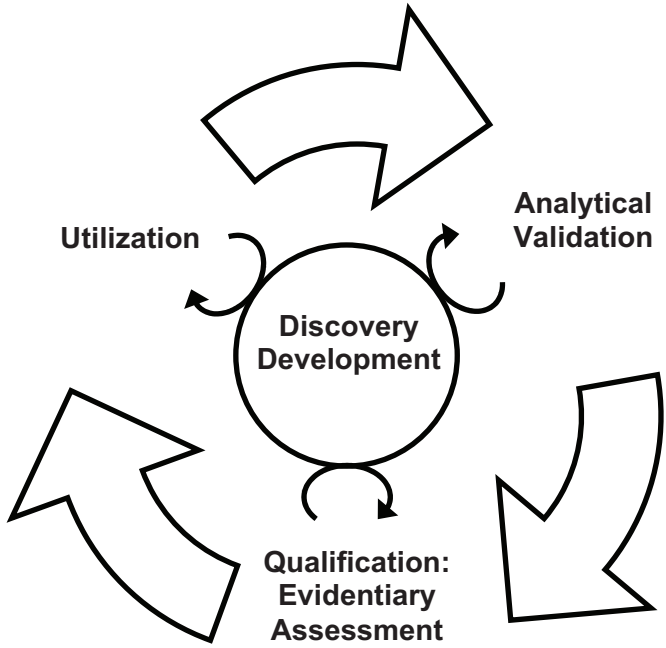

FIGURE 3-1 The steps of the evaluation framework are interdependent. While a validated test is required before qualification and utilization can be completed, biomarker uses inform test development, and the evidence suggests possible biomarker uses. In addition, the circle in the center signifies ongoing processes that should continually inform each step in the biomarker evaluation process.

panels of biomarkers in addition to single biomarkers and for both circulating and imaging biomarkers. While the report provides case studies of individual biomarkers, the committee concluded that sets of biomarkers need to be qualified using the same process. In some cases, individual biomarkers within the same set may need to be qualified individually.

This chapter explores the rationale behind the committee's decision to separate evaluation into three interrelated steps before providing an in-depth examination of each step. This conceptual framework is meant to provide a clear, adaptable platform for statistically sound, evidence-based biomarker evaluation.

THE RATIONALE FOR AN INTERRELATED, THREE-STEP PROCESS

Recommendation 1:
The biomarker evaluation process should consist of the following three steps:

1a. Analytical validation: analyses of available evidence on the analytical performance of an assay;

1b. Qualification: assessment of available evidence on associations between the biomarker and disease states, including data showing effects of interventions on both the biomarker and clinical outcomes; and

1c. Utilization: contextual analysis based on the specific use proposed and the applicability of available evidence to this use. This includes a determination of whether the validation and qualification conducted provide sufficient support for the use proposed.

The committee recognizes that including analytical validation in the evaluation framework and separating the evidentiary assessment from the utilization analysis is a departure from many previous attempts to develop biomarker evaluation systems, but found that these processes, although distinct, are interwoven in such a way that it is impossible to responsibly consider one without also considering the others. Although biomarker analytical validation and biomarker qualification will often be considered together (the statistical linkages of disease, biomarker, and drugs can depend on the analytical soundness of a biomarker assay) and have been used synonymously in the past (Biomarkers Definitions Working Group, 2001), differentiating these processes is important (Lee et al., 2006). A National Institutes of Health working group recommended the term "validation" be used for analytical methods (Biomarkers Definitions Working Group, 2001). The American Association of Pharmaceutical Scientists (AAPS), the Pharmaceutical Research and Manufacturers of America, and the Biomarkers Consortium, among other organizations, have worked to reinforce the distinction between analytical validation and qualification (Lee et al., 2005; Wagner, 2002). As discussed below, analytical validation is the process of assessing how well an assay quantitates a biomarker of interest; qualification is the evidentiary and statistical process linking a biomarker with biological processes and clinical endpoints (Biomarkers Definitions Working Group, 2001). The committee determined that qualification could be further separated into evidentiary assessment and utilization analysis, so that the different investigative and analytical processes required to evaluate evidence and contexts of use

are distinct. Details regarding methods for the gathering of evidence are discussed in the section on Recommendation 2.

It is important to emphasize the necessity of evaluating data relating to adverse events and unintended effects of biomarker use. In every step, the proposed use and its context are critical. For drug development and other medical uses, this entails a risk–benefit analysis, which weighs evidence supporting biomarker use against known inaccuracies and gaps in knowledge that present the possibility of error. For foods and supplements, this entails an analysis of the potential modifying matrix effects of the food or supplement that serves as the delivery vehicle and the dietary patterns associated with consumption of the nutrient or food substance.

The committee understands that a biomarker evaluation checklist of criteria to fulfill for given purposes would be more straightforward to use. But, given the complexities of biomarker utilization, the risks involved with their use, and the evolving nature of science and technology, a checklist-based approach was deemed to be infeasible. First, because any attempts to evaluate a biomarker must consider the context of and purpose for use of the biomarker, scientific and medical judgment plays a role in decision making. Because the purpose and context in each evaluation are unique, there are no precisely relevant past data to consult for guidance. Also, decisions made during the evaluation process are based on probabilistic rather than deterministic reasoning. Probabilistic reasoning emphasizes epidemiological and statistical relationships and acknowledges that the biology is not fully understood. Both statistical methods and decision analysis may be important tools for biomarker evaluations. Both of these were discussed in Chapter 2.

Despite these important caveats, a nuanced understanding of the strength of a biomarker is necessary to develop an evidence-based understanding of whether the biomarker is fit for its proposed purpose and context of use. The committee acknowledges that decisions resulting from the evaluation of a biomarker are dependent on the purposes for which the biomarker will be used. Although some have supported the idea of biomarker evaluations that can be viewed as general and definitive for any proposed purpose or context of use, the committee has determined that there has been no example of this so far and it does not expect to witness one in the future.

The committee recognizes that this approach will require some additional financial and human resources at the Food and Drug Administration (FDA), as was suggested in the Institute of Medicine (IOM) report *The Future of Drug Safety* and is discussed further in Chapter 5 (2007). However, the process fits well with the mechanisms that the FDA already uses to seek external advice (e.g., the scientific advisory committees). Also, this process would represent a modest investment compared to its

potentially broad benefits to society by ensuring a stronger evidence base underpinning FDA decisions. Benefit to the FDA itself and its commercial users may also be realized through more consistent and transparent expectations.

Analytical Validation

As previously defined, the term "biomarker" refers to a characteristic that is reliably and accurately measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention (Biomarkers Definitions Working Group, 2001). Thus, measurement itself must be an explicit component of any discussion of biomarker evaluation because it establishes the scientific basis and availability of experimental data that support or refute the context for qualification of a biomarker and its proposed application (Goodsaid and Frueh, 2007). The committee finds that analytical validation of a relevant biomarker test is a prerequisite for biomarker qualification.

Analytical validation is defined as an assessment of assays and their measurement performance characteristics, determining the range of conditions under which the assays will give reproducible and accurate data. Thus, analytical validation is an assessment of a biomarker test that includes the biomarker's measurability and the test's sensitivity for the biomarker, biomarker specificity, reliability, and lab-to-lab reproducibility. The terminology used in the recommendation, analytical performance, is not meant to describe how well a biomarker correlates with the clinical outcomes of interest. Instead, analytical validation of an assay includes the biomarker's limit of detection, limit of quantitation, reference (normal) value cutoff concentration, and the total imprecision at the cutoff concentration. These specifications must be determined and met before data based on its use can be relevant in the qualification steps of biomarker evaluation. To ensure comparison across multiple laboratories and clinical settings, appropriate standards for ensuring quality and reproducibility need to be made available. Additionally, understanding the difference between individual assays is important to interpreting the findings of different studies monitored using different assays (Apple et al., 2007). For biomarkers used solely in laboratory testing, it would be beneficial to assess the ability to compare data from different assay platforms as much as possible and needed.

Though key guidelines and regulations have molded approaches to assay validation (Swanson, 2002), biomarker validation is distinct from pharmacokinetic validation and routine laboratory validation; however, an agreement for a uniform set of criteria for biomarker assay validation

has not been reached. Method validation requirements for assays that support pharmacokinetic studies have been the subject of intense interest, and the FDA has issued guidance for industry on bioanalytical method validation (CDER, 2001).[1] This guidance, though, is directed at validation of assays looking at metabolism of conventional small-molecule drugs and is not directly related to the validation of assays for biomarkers for many other uses. Similarly, biomarker validation is, in many ways, different from routine laboratory validation. Laboratories that perform testing to support human diagnostics and health care are regulated by the *Clinical Laboratory Improvement Amendments*, or CLIA (Centers for Medicare & Medicaid Services), and accrediting organizations such as the College of American Pathologists (Swanson, 2002).

Because of the diverse purposes of biomarker research and the various locations in which these assays are performed (routine and novel biomarker assays are performed in both bioanalytical and clinical laboratories; novel biomarker assays are also performed in specialized laboratories), neither the FDA regulations nor the CLIA guidelines fully address all possible study objectives. Differences between biomarker assays and those of drug bioanalysis and diagnostics are described in detail in Table 1 of Lee et al. (2006), highlighting some of the unique validation challenges related to biomarker assays (Lee et al., 2006).

In the absence of uniform criteria for the validation of biomarker assays, analytical qualities and clinical performance of assays cannot be objectively evaluated (Apple et al., 2007). To address these challenges, the AAPS and Clinical Ligand Assay Society cosponsored a Biomarker Method Validation Workshop in October 2003 (Lee et al., 2005). It resulted in a validation approach for laboratory biomarker assays in support of drug development. This validation approach, though, was focused primarily on ligand-binding methods to gauge biomarkers measured ex vivo from body fluids and tissues (Lee et al., 2006).

Additionally, the International Federation of Clinical Chemistry and Laboratory Medicine Committee on Standardization of Markers of Cardiac Damage has recommended analytical and preanalytical quality specifications for a variety of assays, including those for natriuretic peptides and troponin assays (Apple et al., 2005, 2007; Writing Group Members et al., 2008). These guidelines were developed to guide both clinical and commercial laboratories that use the assays with the goal of establishing uniform criteria (Apple et al., 2007). The standardization and harmonization of biomarker assays is challenging due to the various analytical and biological factors that influence measurement (Swanson, 2002). By defini-

[1] *Good Laboratory Practice for Nonclinical Laboratory Studies.* 2001. Code of Federal Regulations, Title 21, Vol. 1, 21 C.F.R. 58.

tion, biomarkers are dynamic and responsive to changes in the disease process, pharmacological intervention, and environment (Fraser, 2001; Ricos et al., 1999). For example, variability in biomarker level is affected by biology (e.g., gender, age, posture, diet), sample type (e.g., blood, urine), sample collection (e.g., transport and storage conditions, collection technique), and analytical factors (e.g., pipetting precision, antibody specificity) (Swanson, 2002). Sources of variability in biomarker measurements are summarized in Table 3-1. Though these background fluctuations affect

TABLE 3-1 Sources of Variability in Biomarker Measurements

Preanalytical Sources of Variability		Analytical Sources of Variability
Biological	Sample Collection	
Sociodemographics (including age and gender)	Mislabeling	Purity of reference standards
Posture	Duration of tourniquet application	Lot-to-lot variation in reagents
Exercise	Strength of collection vacuum	Antibody crossreactivity
Meals/fasting status	Size of needle gauge	Loss during extraction
Diet	Dead volume in catheters/collection tubes	Mislabeling of processing tubes
Diurnal biorhythm	Anticoagulants	Pre-assay incubation time and temperature
Seasonal biorhythm	Local effects of indwelling catheter	Pre-assay amplifications
Concurrent diseases	Time and temperature prior to centrifugation	Chemical interference by endogenous compounds
Concurrent medications	Centrifugation speed, duration, temperature	Chemical interference by drugs
Overall health/ preexisting disease	Evaporation/biomarker volatility	Analyte or reagent instability in light
Gastrointestinal motility	Preservatives/biomarker instability	Time between intermediate steps
Anesthesia/surgical intervention	Storage temperature	Fluctuations in instrument performance
Stress	Transport temperature	Correction for baseline/ background levels
Pregnancy	Completeness of urine collection	Post-run calculation errors
Menstrual cycle	Hemolysis	Matrix effects
Dehydration	Effect of glass and plastic collection tubes	Reproducibility of sample
Kidney function	Exposure to light	
Body composition (obesity)	Type of sample	
	Time of clotting	

SOURCE: Adapted from Swanson (2002). Copyright 2010, reprinted with permission from IOS Press.

both the sensitivity and the specificity of biomarker measurements, and though it may not be possible to establish absolute accuracy, relative accuracy can be informative and the sources of variability can be understood and controlled, allowing for the delivery of high-quality assay results (Lee et al., 2006; Swanson, 2002).

Implementation of biomarker validation therefore requires both understanding and control of the various sources of variability in assay performance (Kristiansen, 2001). Results from biomarker assays are valid only if sample integrity is maintained from sample collection through analysis. It is important to devise standard protocols for sample collection, processing, and storage to achieve uniformity (Lee et al., 2005, 2006). The committee synthesized a variety of approaches to develop its key elements for biomarker validation. Table 3-2 lists important data for inclusion in package inserts and in peer-reviewed publications for biochemical biomarker assays in the preanalytic characteristics, calibration and standardization criteria, and analytic parameters. Other considerations may be needed for imaging and other types of biomarker assays; this is discussed further in Chapter 4.

Validation of biomarker tests should be done on a test-by-test basis and must then be deemed sufficient for the use proposed in the utilization step (ICH, 1994; Shah et al., 1992). Thus, the rigor of biomarker validation can be correlated with the intended use of the data (Lee et al., 2006). The committee finds that biomarker qualifications are often undermined by insufficiently validated tests, which may lack accuracy, sensitivity, and specificity. Additionally, use of tests after biomarker qualification and test validation depends on operator, reagent, and instrument variability, among other factors. In the case of clinical laboratory assays and reference ranges of common biomarkers, for example, absence of standardization can lead to interpretation mistakes (Rosner et al., 2007; Wu, 2010). The nature of health care is such that patients often use multiple laboratory facilities during the course of care (Wu, 2010). Diagnosis and management depend on the accuracy of testing across laboratories (Rosner et al., 2007). Therefore, proper standards and controls are necessary to ensure consistent delivery of high-quality biomarker data and the validation of biomarker tests prior to biomarker qualification. Box 3-2 introduces the case study exemplifying the issues found in analytical validation. Further detail can be found in Chapter 4.

Qualification

The second step of the committee's evaluation framework is a factual description of the levels and types of available evidence. This objective analysis is a reproducible, systematic assembly and review of the evi-

TABLE 3-2 Information Needed for Package Inserts and Peer-Reviewed Publications Describing Biomarker Assays

Preanalytic	Calibration/ Standardization	Analytic
Sample handling	A low-level quality control (QC) sample with concentration close to reference value to monitor assay bias at cutoff	For antibody assays, identification of antibody recognition epitopes
Effect of storage time and temperature	A negative QC sample to monitor baseline drift	For activity assays and immunoassays, identification of limiting substrates
Influence of different anticoagulants (type and concentration) for plasma and whole blood measurements	Calibration frequency to be determined based on the imprecision and drift characteristics of the assay	Linearity of signal
Influence of gel separator tubes	Calibration using defined biomarker calibrators to accommodate any subtle changes in assay calibration curve	Reactivity to various plasma biomarker forms (degree of equimolarity)
Time and speed (relative centrifugal force) and temperature of sample centrifugation with the effects of various methods for tube filling, mixing, and centrifugation	Defined limits for the zero calibrator's reaction units	Cross-reactivity with other related proteins in complex matrix (normal and disease)
		Identification of interferences from hemolysis, bilirubin, and lipemia, and potential interferences from heterophile antibodies, rheumatoid factors, and human antianimal antibodies and autoantibodies (neither of which are currently commercially available)
		Dilution response (i.e., linearity, recovery) over time and sites
		Assay limit of blank, limit of detection, and limit of quantitation
		Decision limits and precision at relevant concentrations
		Method comparison data, in particular if manufacturers offer both central laboratory and point-of-care assays
		Establishment of the decision limit of the distribution of healthy subject reference values

BOX 3-2
Tumor Size and Analytical Validation (Recommendation 1a)

Tumor size is a variously defined biomarker of efficacy of cancer therapeutics using tumor diameter, tumor volume, or tumor mass, as measured by a variety of platforms and techniques, including magnetic resonance (MR), computed tomography (CT), and positron emission tomography (PET). Different contrast agents and different protocols may be used, all of which affect the precision of measurement. Measurement precision is also affected by patient characteristics. Each protocol, which may also vary by tumor location, should undergo independent validation. There is a great deal of variability in the levels of evidence to support validation for different protocols; thus, analytical validation is complicated by multiple imaging platforms and other assay performance issues. The disparity in evidence impacts the interpretation and generalizability of these imaging endpoints.

Assuming that at least one test is determined to be adequately validated, data collected for the qualification step have shown that tumor size may not always be linked to clinical benefit although tolerance for uncertainty of clinical benefit has been justified by the seriousness of cancer.

For utilization, in 1992, the Food and Drug Administration started granting accelerated approval for drugs that are effective against serious diseases based on surrogate endpoints. Accelerated approvals for anticancer drugs or biologics have been granted on the basis of endpoints such as overall response rate, time to progression, or disease-free survival. Of those granted approval between 1992 and 2004, only about one-quarter have been converted to regular approval (i.e., demonstrating an effect on survival) (Lathia et al., 2009). All of them remain on the market. Concern exists that clinical benefit may be neglected in regulating this type of approval (Fleming, 2005). Tumor size is discussed in greater detail in the full case study found in Chapter 4.

dence. Users of the evaluation framework will need to identify appropriate methods for gathering the evidence for this step. This is discussed with respect to the FDA in the section on recommendation 2. Fulfilling the qualification step requires: (1) evaluating the nature and strength of evidence about whether the biomarker is on a causal pathway in the disease pathogenesis, and (2) gathering available evidence showing that interventions targeting the biomarker in question impact the clinical endpoints of interest. If the biomarker–clinical endpoint relationship persists over multiple interventions, it is thought to be more generalizable.

It is important to note that although this is an objective, evidence-based assessment, the type of reasoning that may be used in this step is still probabilistic rather than deterministic. While deterministic reasoning ultimately means that every contributing factor to the biomarker–intervention–clinical endpoint link is defined and understood, probabi-

listic reasoning emphasizes epidemiological and statistical relationships, acknowledging that all contributing factors are generally not fully understood. Because this is almost always the case, clinical outcomes are fundamentally random in nature, requiring probabilistic reasoning to inform rational decision making. Thus, biomarker evidence allows for inferences, but rarely allows for certainties.

Evidence for a Link Between the Biomarker, the Disease Pathway, and the Clinical Endpoint: Hill's Criteria

For the first part of the qualification step, evaluating the strength of evidence regarding the disease pathway can be done, in part, by using concepts described by Hill's criteria (1965). Hill's criteria were discussed in detail in Chapter 2; they evaluate characteristics such as strength of association, biological plausibility, and consistency, among others (see Box 2-2 and supporting text in the previous chapter) (Williams et al., 2006). Understanding the biology behind a biomarker is an important source of information on a biomarker's relevance, specificity, and robustness (Koulman et al., 2009). However, biomarkers indicating differences between healthy and sick individuals may relate to consequences rather than the causes of the underlying disease pathology; as a result, these differences may not have predictive value (Koulman et al., 2009). As a result, these biomarkers need to undergo a rigorous multistep qualification process in order to become diagnostic tools (Koulman et al., 2009).

Given that biomarkers are "indicators"—in that they are not necessarily causal—and that an abnormal value or a gradient in level over time is not necessarily informative or predictive depending on the clinical situation, the committee instead used these criteria as a structure for assessing the prognostic value of the biomarker for the clinical outcomes of interest. Depending on the situation, not all of the criteria must be fulfilled; temporality and strength of association are generally necessary, however.

Different study designs have advantages and disadvantages. Prospective or cohort studies allow researchers to define study populations based on some relevant characteristic(s) in advance (e.g., level of a biomarker), then follow the development of health outcomes over time. This process is the most accurate and inclusive of possible outcomes, but it is slower than some other designs. On the other hand, cross-sectional studies define a population of interest and then collect data on both the characteristics of interest (e.g., level of a biomarker) and the health outcomes of interest simultaneously. Although these studies are faster, they have limitations. Cross-sectional designs do not allow for causal inferences to be made since biomarker–disease measurements occur simultaneously. Also, patients who died or experienced clinical outcomes that made them unavailable

for measurement would not be reflected in a cross-sectional population, leading to significant risk of incorrect conclusions. This is thought to be a reason that lower low-density lipoprotein (LDL) levels have been found to be associated with higher risk of death in patients after cardiac catheterization, even though lowering LDL cholesterol with statins has a large benefit in these same patients (Califf et al., 1992). Therefore it is important to consider the quality and strength of the data when conducting biomarker evaluation.

Evidence That Interventions Impacting the Biomarker Impact the Clinical Endpoint

For the second part of qualification—for surrogate endpoints, that is—prognostic value is a necessary but not sufficient criterion for the evaluation. Interventions targeting the biomarker in question should impact the clinical endpoints of interest. Although laboratory or preclinical data may indicate the effect of interventions on the biomarker and correspond to the effect on clinical outcome, robust, adequately controlled clinical study data using clinical endpoints (i.e., phase III data or equivalent studies) are necessary. Observational data in human populations and preliminary clinical data (e.g., phase I or II data) are considered, but are not sufficient to fully qualify a biomarker as a surrogate endpoint at this stage of evaluation. An informative evidence-based approach to qualification of a surrogate endpoint may be based on an overview analysis of multiple randomized trials, where the relationship of intervention's effect on the biomarker is plotted against the intervention's effect on the true clinical endpoint. Examples of this include an assessment of progression-free survival as a surrogate endpoint for overall survival for adjuvant treatment of colorectal cancer (Fleming, 2005; Sargent, 2004) and the assessment of blood pressure as a surrogate endpoint for cardiovascular risk (Staessen et al., 2003).

In biologic systems, a given intervention can exert multiple different, even contradictory, actions. There are challenges to determining what clinical trial data are necessary to document the value of interventions to target the specific clinical endpoint and predict benefit and harm. For example, postmenopausal hormone replacement therapy (HRT) was thought to protect women from cardiovascular disease based on both observational, epidemiologic data and the apparent beneficial effects of estrogen on lipoproteins and other cardiovascular disease biomarkers. However, several important inflammatory biomarkers (adhesion molecules) and prothrombotic biomarkers were not measured in the early studies. After several clinical trials, HRT was discovered to raise mortality from cardiovascular events and have other adverse unexpected effects.

For this reason, the committee recommends analyzing multiple mechanistic pathways leading to the same outcome when evaluating biomarkers.

Additionally, interventions may impact populations differently. The clinical consequences of an intervention may differ from a healthy population to a population with extensive comorbidities. A biomarker must not merely show a difference between healthy controls and individuals with a certain disease; the control group must also include subjects with other pathophysiologies in order to ensure the biomarker data show a distinctive difference between controls and individuals with the disease (Koulman et al., 2009). In the description of evidence about the biomarker, populations and conditions to which the assessment applies need to be articulated so they can be considered in the utilization step of the biomarker evaluation framework.

Box 3-3 introduces the case study exemplifying issues surrounding biomarker qualification. This example is discussed in Chapter 4.

BOX 3-3
CRP, Inflammatory Markers, and Qualification (Recommendation 1b)

As the scientific understanding of atherosclerosis has evolved to include inflammation's role in the disease process, researchers have sought inflammatory biomarkers. The one most extensively studied is C-reactive protein (CRP). In observational studies, CRP is an independent predictor of future vascular events, including myocardial infarction, ischemic stroke, peripheral vascular disease, and vascular death. In spite of CRP's utility in cardiovascular risk prediction, its normal function and role in cardiovascular disease remains uncertain. The lack of understanding of CRP's biological role in human physiology has elicited controversy over assertions of CRP's causal role in cardiovascular disease.

CRP satisfies the first step of the evaluation framework, analytic validity: CRP is easily measured by standardized high-sensitivity immunoassays and has negligible diurnal variation, does not depend on food intake, and has a long half-life. For qualification, CRP has prognostic value: a number of population cohorts have shown CRP to predict future vascular risk, though with some caveats (these are explained in Chapter 4). The evidence supporting CRP's biologic association with cardiovascular disease is weak, and more research is needed to clarify determinants of CRP variation and utility in diverse populations. Although several interventions are known to lower CRP, it is unclear whether there is consistency of correlation between the effects of different interventions on CRP and clinical outcomes. Based on these findings, in the utilization step, CRP would not qualify for the context of use of surrogacy, but it may be used in risk prediction in certain populations. This matter is discussed in greater detail in the full case study found in Chapter 4.

Utilization

The third step of the committee's biomarker evaluation framework is a contextual analysis of the available evidence about a biomarker with regard to the proposed use of the biomarker. These evaluations should take place on a strictly designated fit-for-purpose basis, with consideration for the context of use, as knowledge and technology continually evolve. Defining the context of use requires explicit articulation of the populations and conditions for use to which the assessment applies. For surrogate endpoints, idealized statistical requirements are rarely or never achievable; subjective assessment is necessary to determine when surrogate endpoints can be used. This variability between evaluations can be minimized by consistently evaluating the critical and important factors, including risk assessment, as described by the committee.

The utilization step can be divided into several components. As discussed in Chapter 2 of this report, biomarkers have a multitude of uses in both clinical care and drug development, including for risk stratification, prevention, screening, diagnosis, prognosis, patient selection, and pharmacodynamics (see Table 2-1). In drug development, for example, there is a continuum of uses from early trials on one end to surrogate endpoints on the other end. A determination of the general category of use for which the biomarker is intended is necessary to inform the evaluation process as to whether analytical validation and qualification data are appropriate for that use, particularly when it may be involved in future regulatory or policy decisions. The list of uses is further expanded when biomarkers are discussed in the arenas of medical devices, biologics, and nutrients and foods. This determination can therefore be understood as a necessary first component in the utilization analysis step of biomarker evaluation.

The second component in the utilization analysis is consideration of factors related to defining the context for which a biomarker should be qualified. Generally, the earlier in the development of an intervention, the more flexibility there is in using a biomarker. The committee evaluated a multitude of factors, including prevalence of the disease, risks associated with the intervention, and concurrent and prior treatment, to develop its criteria. The exhaustive list of factors was synthesized into a concise list of Critical and Important Factors (see Table 3-3). The recommendations are meant to provide general guidelines that could be adapted for all uses of biomarkers that result in clinical, product, or claim development, or regulatory decisions, whether for drugs, biologics, or device development; for relationships between diet or nutrients and disease; or for public health monitoring and interventions. Thus, the criteria are broad in scope.

One of the principal considerations in biomarker evaluation is whether the biomarker is being used as a surrogate endpoint. If it is, the standards of evaluation are more rigidly defined. This scenario is discussed further

TABLE 3-3 Utilization: Critical and Important Factors for Consideration (Recommendation 1c)

	Factors	Rationale
Critical	1. Is the biomarker being used as a surrogate?	If the biomarker is used as a surrogate, enhanced scrutiny would be necessary.
	2. What is the prevalence of the disease? What are the morbidities and mortalities associated with this disease?	A highly prevalent or serious disease might have a lower threshold for use of biomarkers in clinical and regulatory decisions.
	3. What are the risks and benefits associated with the intervention? Has due attention been paid to both safety and efficacy?	The benefits of the intervention must be weighed against the risks of biomarker failure to define a range of tolerable biomarker performance for each specific biomarker (Williams et al., 2006).
	4. What are the advantages and disadvantages associated with use of the biomarker when compared with the best available alternative? How does the biomarker benefit management and outcomes?	The evaluation may proceed differently depending upon whether a variety of valid treatment options are available compared to if no treatments have yet been developed, for example.
	5. Is the biomarker for drugs, biologics, or device development; for relationships between diet or nutrients and disease; or for public health monitoring and interventions?	While the highest level of scientific rigor is needed in biomarker evaluations for all uses, each category of use has different risks and regulatory frameworks, which carry implications for appropriate evidence thresholds and requirements for biomarker use.
Important	6. What is the biomarker's purpose with respect to phase of development in clinical trials?	For biomarkers that are likely to be used in a regulatory submission or as evidence supporting statements regulated by the FDA, consideration should be given to the need for additional data collection.
	7. Is the biomarker for primary or secondary disease prevention?	Biomarkers used for these purposes carry especially high risk and should be evaluated with this consideration in mind.

in the next section. For all biomarkers, including surrogate endpoints, the prevalence of the disease and the morbidities and mortalities associated with the disease are important contextual considerations. For example, in general, use of an intervention meant for primary prevention will have an extremely low tolerance for risk. Within this minimal tolerance, however, for risk reduction of a very common, serious chronic disease, more risk may be tolerated than for an intervention intended to prevent a less common or less serious disease. Likewise, an intervention meant to treat a rare but life-threatening disease may permit more tolerance of risk than an intervention meant to treat a more common but less serious disease. So, it may be easier to defend use of a surrogate endpoint for trials of rare and life-threatening diseases than for trials of primary prevention interventions for common but less serious or life-threatening diseases.

The safety and efficacy of biomarker use can be thought of in conjunction with the risks and benefits associated with the intervention targeting the biomarker. The benefits of the intervention must be weighed against the risk of biomarker failure to define a range of tolerable biomarker performance for each specific biomarker (Williams et al., 2006). Subjectivity can be minimized by thinking of biomarker utility as analogous to risk assessment, as discussed by Williams and colleagues (2006). In Williams's proposed framework, the generation of knowledge links specific risk agents with uncertain, but possible, outcomes. Thus, a key factor is the perceived consequence that would result if the biomarker were to fail. Although quantitative information related to the degree and frequency of failure may be unavailable, the seriousness of this failure should be a factor in evaluation (Williams et al., 2006).

The extent to which a given surrogate endpoint can be a target of a therapeutic intervention depends on a patient's or population's specific constellation of risk factors, relative to the multiple components of risk found in the population as a whole. It is important to determine the mechanism dominating the clinical effect so that interventions most likely to affect that mechanism can be selected for particular patients or populations. For example, for patients with familial hypercholesterolemia, high LDL cholesterol (LDL-C) will likely be their most important cardiovascular risk factor; interventions that target LDL-C may be justified even when there is a lower level of supporting evidence (i.e., use of interventions approved on the basis of surrogate endpoint data). For the general population, on the other hand, where competing cardiovascular risks are from high LDL-C, hypertension, inflammation, smoking, and other dyslipidemias, the successful use of LDL-C as a surrogate endpoint for cardiovascular risk is less assured. Hence, better evidence is needed on the connection of LDL-C–lowering interventions and clinical outcomes

before use of those interventions can be recommended in the general population.

The committee finds that the hazards of making the wrong decision regarding a biomarker's qualification is a critical factor in the decision-making process. Although the opportunity cost (i.e., the loss of the benefits of the next best alternative decision) differs depending on the stakeholders, the subjectivity of this consideration can be minimized. When a choice is made, the opportunity cost is the benefit that would have occurred had the second best option been chosen instead. To illustrate opportunity cost, someone choosing a breakfast food in the absence of health advice might choose a food high in calories and saturated fat or they might choose a bowl of healthy cereal with skim milk and an apple. The opportunity cost for one of these decisions over the other is the potential health benefit advantage or money saved that would have occurred should the other option have been chosen. In a related example, a box of cereal carries a claim recommending its healthy characteristics. In this case, an individual may choose the cereal with the healthy claim over a similar, cheaper cereal, or choose it over an unhealthy option. The opportunity cost would be the money that would have been saved by choosing the cheaper cereal, but the choice may also have prevented the individual from choosing an unhealthy breakfast. For the cereal manufacturer, the opportunity cost of not carrying the healthy claim would be the lost profits of more individuals choosing the manufacturer's cereal, whereas the opportunity cost of carrying the healthy claim would be the money that could have been saved by not developing the claim, printing the new packaging, or carrying the legal liability of the claim.

As with considerations of competing risks, knowledge of the concurrent and prior treatments used in treating an individual patient or a patient population plays a role in contexts of use for which a biomarker may be qualified. The evaluation may proceed differently depending upon whether a variety of treatment options are available compared to if no treatments have yet been developed, for example. The committee believes it is important to value the costs of denial of an intervention to patients who would benefit.

The committee did not explicitly include analysis of a biomarker test's or intervention's cost effectiveness in the evaluation framework. Cost effectiveness is important for a subset of biomarker uses, particularly those involving changing the clinical practice of medicine. In such situations, evaluators may wish to include analysis of cost effectiveness of interventions in the utilization step. A great deal of research has been done on how to conduct such studies, although the committee cautions that definitive estimates of costs can be made only after clinical outcomes are measured.

BOX 3-4
Troponin and Utilization (Recommendation 1c)

Use of troponin as a biomarker in acute settings is ubiquitous as a method to diagnose myocardial infarction (MI). MI causes cardiac muscle damage that results in a rise in troponin concentrations. Its use in chronic settings is more recent, and relies on developing high-sensitivity assays that still require validation. However, the criteria for such a validation are advanced compared to the current regulatory standards. Troponin can be elevated in patients who may suffer from a variety of chronic heart conditions, inflammatory conditions, side effects from drugs, or organ failures. These assays have not yet shown analytical validation. But, should one or several of the assays eventually show adequate sensitivity, specificity, and reproducibility, then the biomarker can be advanced to the qualification step. In qualification, it is apparent that clinical data from several different trials (Gupta and de Lemos, 2007; NACB Writing Group Members et al., 2007) show increased risk of mortality in individuals with elevated troponin levels. However, although there is evidence that prevention of MI reduces death rates, there is no evidence that using an intervention to decrease troponin levels rather than preventing the event in totality improves mortality risk. Finally, although use of troponin as a biomarker in phase I studies to indicate cardiac safety problems with tested drugs or to collect further information about the valuable applications of this biomarker is justified and valuable, use of troponin levels as a surrogate endpoint for interventions is not justified due to a dearth of evidence. This matter is discussed in greater detail in the full case study found in Chapter 4.

Box 3-4 summarizes the case study for the biomarker troponin, which illustrates some of the judgments that can be made in the utilization step of biomarker evaluation. This case study is discussed in further detail in Chapter 4.

Evaluation of a Biomarker as a Surrogate Endpoint

In the case of chronic disease, where there are multiple pathogenetic pathways leading to development of clinical outcomes and multiple manifestations of disease, the probabilistic nature of predictions made using biomarker data means that no biomarker can give absolute certainty of an event's future occurrence nor absolute certainty of the timing of the predicted event. Nonetheless, there are situations in which use of a biomarker as a surrogate endpoint in situations with regulatory impact may be supported, such as in situations where the need for interventions is urgent or where studies including clinical endpoints are not feasible because of technical or ethical reasons. Again, this is not meant to discourage use of biomarkers in product development; biomarkers play an important

role in research and decision making. Situations with regulatory impact are defined in the section on Recommendation 2. Finally, it is essential to remember that the information that an individual surrogate endpoint or clinical endpoint can give is inherently limited; as a result, it is important to emphasize the need to evaluate data relating to adverse events and unintended effects of biomarker use.

The committee does not intend to imply that selection of endpoints for clinical trials would be simple or risk free if investigators were simply to avoid surrogate endpoints. Clinical and surrogate endpoints have been defined in a way that may imply a clear distinction between the two, in that clinical endpoints typically reflect patient experience and surrogate endpoints do not. However, there is discussion surrounding this issue, which illustrates the scientific complexity of the distinction between clinical and surrogate endpoints. Some clinical endpoints have many similarities with biomarkers, and can be thought of as a step removed from patient experience, and therefore subject to similar potential failings as surrogate endpoints (i.e., pain scales). Some surrogate endpoints are highly robust (i.e., HIV-1 RNA for particular classes of viral-suppressing drugs). However, even these endpoints require an understanding of unrelated effects, the magnitude and duration of target effects, and optimization of use (such as timing related to initiation of viral-suppressing drugs). Clinical endpoints share many features of biomarkers, such as the need for analytical validation, but they differ from biomarkers in that clinical endpoints address how a patient feels, functions, or survives and also commonly utilize multiple diagnostic criteria. Nonetheless, the committee recognizes that selection of clinical endpoints is beyond the scope of this report. There are many important interests at stake in this discussion and some issues, such as the best way to choose endpoints for trials, may be context specific. In such settings, stakeholders such as industry, the public as represented by government and community representatives, and academic researchers, may benefit from convening to discuss these issues.

Utilization aims to establish whether the biomarker is being used as a surrogate endpoint; the prevalence, morbidity, and mortality of the disease; the risks and benefits associated with the intervention; and opportunity cost, among other factors. Box 3-5 introduces the case study on surrogate endpoint status of LDL and high-density lipoprotein (HDL) cholesterol.

APPLICATION OF THE EVALUATION FRAMEWORK

Recommendation 2:

2a. For biomarkers with regulatory impact, the FDA should convene expert panels to evaluate biomarkers and biomarker tests.

2b. Initial evaluation of analytical validation and qualification should be conducted separately from a particular context of use.

2c. The expert panels should reevaluate analytical validation, qualification, and utilization on a continual and a case-by-case basis.

Recommendation 2 provides further guidance on the application of the framework to uses of biomarkers that have regulatory impact. Specifically omitted from this recommendation are biomarker discovery activities and biomarkers for use in drug discovery, development, and other preclinical uses. This decision was made based on the sheer volume of

BOX 3-5
LDL and HDL Cholesterol and Surrogacy (Recommendation 1c)

Low-density lipoprotein cholesterol (LDL-C) concentration is considered as a qualified surrogate endpoint for cardiovascular disease for both food-related disease claims and drugs. It is often viewed as the benchmark biomarker (Couzin, 2008; Rasnake et al., 2007). Thus, an examination of the evaluation of LDL-C not only highlights the strengths of the biomarker itself, but also the ways in which even qualified biomarkers face contextual caveats. The evidence supporting this biomarker rests almost entirely on the measurement of LDL-C even though it is only one part of the lipoprotein particle. Both apolipoprotein B and the quantity and the composition of LDL particles themselves have potential to be more indicative of cardiovascular disease risk than LDL-C for some populations (Berneis and Krauss, 2002; Rizzo and Berneis, 2007; Tardif et al., 2006), showing that even for qualified biomarkers, developing standard measures is an ongoing process.

The strength of LDL-C as a surrogate endpoint is not absolute due to the heterogeneity of cardiovascular disease processes, the heterogeneity of LDL-lowering drug as well as food effects, and the heterogeneity of LDL particles themselves. Because cardiovascular disease is a multifactorial chronic disease, a single component of the disease (e.g., LDL-C) cannot fully account for all the variability that leads to a particular outcome (Libby and Theroux, 2005; Tardif et al., 2006). The C-reactive protein case study suggests that inflammation, for example, may also affect the cardiovascular disease pathway. Furthermore, sociodemographic factors have been shown to complicate these already complex disease dynamics; as a result, lowering LDL-C can never be assured to be a "perfect" indicator across all population groups or all interventions.

Interventions to address a multifactorial disease introduce potentially unforeseen effects, particularly when the causal disease pathways, the mechanisms of action of the intervention, and the biochemical characteristics and function of the biomarker itself are not fully understood. High-density lipoprotein (HDL) does not qualify as a generic surrogate endpoint because these characteristics, particularly the latter, introduce high levels of variability. Furthermore, evidence is weak that elevation of HDL from therapeutics decreases cardiovascular disease risk. LDL and HDL are discussed in greater detail in the full case study found in Chapter 4.

newly discovered biomarkers and the low probability that any given new biomarker will see application in either a regulatory submission for a new product or in a clinical setting. The committee sought ways to achieve a rigorous evaluation framework without stifling innovation. Experts qualified by experience and training are needed to conduct the evaluation reviews, focusing on utilization, as case-by-case analyses are the only way to ensure proper use of biomarkers given the state of the science. The committee has sought ways to support the FDA and other federal agencies in their important work to maximize public health and to regulate the food and drug industry. Maximizing public health means not only protecting the public from inconclusive or fraudulent claims about a food's or drug's benefits on health, but also allowing for rapid access to effective drugs and health-related information. Routes to evidence-based regulation need to be sought that will account for continued innovation and development of products and strategies to improve human health.

Situations having regulatory impact were defined by the committee as follows: circumstances where biomarker data will be submitted, is anticipated to be submitted, or may be requested for submission (as in the case of verification of certain claims on foods or supplements) to the FDA for a regulatory purpose. This definition allows for biomarker discovery and early product discovery activities without convening expert committees. The committee also considered situations in which generally accepted criteria for approval of drugs and other interventions cannot be followed due to insufficient numbers of patients, such as for development of interventions for rare diseases. In the case of rare diseases, product applications are submitted through the FDA's Office of Orphan Products Development (OOPD). The committee suggests that OOPD will need to adapt use of Recommendation 2 to fit with its task.

Direct engagement by the FDA in the process of biomarker evaluation for regulatory decision making may be helpful. Because of the substantial expense, resources, and time that will be needed to qualify new biomarkers, particularly as surrogate endpoints, prospective and specific guidance on the potential or actual acceptance of biomarkers on the part of regulatory agencies for different purposes, and the agencies' regulatory risk tolerance in qualifying biomarkers for each new use, would also be helpful. Ideally, this would lead to an agreement on the weight and specificity of data that would need to be submitted to qualify biomarkers for each purpose under proposed conditions for use. Such an agreement would help to justify the cost and risk of an elaborate biomarker research program in the same way that an end-of-phase II meeting does for phase III drug development. The IOM is currently conducting a study on "Accelerating Rare Diseases Research and Orphan Product Development"; the report will be released in late 2010 or early 2011 (IOM, 2010a).

This organized, large-scale approach to evaluation of a biomarker with regulatory impact requires the convening of an expert panel, similar to an FDA advisory committee, with (1) appropriate expertise, (2) a variety of stakeholders, and (3) attention to conflict of interest. Due to the complexity of data, the need for context-of-use analysis, and the need to deal with sometimes-contradictory evidence, expert input is essential to provide scientific judgment in areas of uncertainty. Experts qualified by experience and training are needed to conduct the evaluation reviews, as case-by-case analyses are the only way to ensure proper use of biomarkers. The same expert panel can discuss all steps in the evaluation framework provided that the panel contains all needed expertise. Panelists should encompass a range of backgrounds, as well as a full range of areas of expertise, including biologists, pharmacologists, clinicians, clinical trialists, and statisticians, as necessary, for decision making. The panelists must be knowledgeable about the biomarker evaluation process and represent a diversity of disciplines and perspectives.

Numerous entities, including the IOM in a 2009 report *Conflict of Interest in Medical Research, Education, and Practice*, have defined a need for attention to conflict of interest in order to protect the integrity of professional judgment. The expert panel for biomarker evaluation should be formulated with due attention to the 2009 IOM recommendations. The biomarker evaluation process inevitably requires judgments to be made by the expert panel; these judgments must be known to be made in good faith and without undue influence. A well-formulated conflict-of-interest policy does not prohibit individual and institutional relationships that might be questioned, but rather, manages these relationships as necessary and required by the policy (IOM, 2009).

As indicated in Figure 3-1, the steps in the recommended framework interact; they are not necessarily separated in time, and conclusions in one step may require revisions or additional work in other steps. For example, as the case study presented in Chapter 4 indicates, tumor size is a biomarker often used for determining efficacy of cancer drugs. Because of inconsistent definitions of tumor size and new findings about the prognostic value of tumor size for specific cancers, among other factors, tumor size is a biomarker that has been, and will continue to be, continually revisited.

Nonetheless, Recommendation 2b states that initial analytical validation and qualification of a biomarker can and should be conducted separately from a particular context of use. The committee understands that no decisions can be made about use of a biomarker without having its use in mind. The committee concluded that it was important to separate the parts of the evaluation framework that have the goal of being objective and those for which subjective judgment is necessary. Analytical valida-

tion and evidentiary qualification were viewed as objective tasks of gathering available evidence, and so they can be conducted separately from a particular context of use. It is also important to discuss briefly, how the committee envisions conduct of the data-gathering process. Data should be gathered from all available sources of evidence. When the evidence is to come entirely from the public domain, it can be gathered according to principles of systematic review (Cochrane Collaboration, 2009; IOM, 2010b). When data not generally publically accessible is made available, such as data owned by companies, for example, then gathering of such data would likely be subject to the same processes as data submission to the FDA for product review.

Evidence evolves even after a biomarker is evaluated; thus, it is imperative that biomarkers be reevaluated periodically so that both the scientific evidence and context-of-use analyses capture the current state of the science. By continual, the committee refers to the need for regular reevaluation on the basis of new scientific developments and data. For instance, continuing with the tumor-size case example, progression of gastrointestinal stromal tumors was found to occur within the original tumor boundaries. Although chemotherapeutic treatment of the tumors may result in decreased cell density and prolonged survival, tumor size (in terms of measurable diameters) was found to generally remain the same. These findings could be cause for reevaluation of the analytical validation step of the biomarker evaluation framework.

Ideally, research findings would dictate the necessity for reevaluation. Post-hoc review should be performed at regular intervals as new information is available to determine how new conclusions should modify the biomarker's qualification and use. When new, potentially relevant evidence related to a biomarker is found, this evidence would be considered to determine the continued appropriate use of the biomarker across a variety of contexts. In practice, however, research efforts are often piecemeal and new findings may not readily be identified as cause for reevaluation of a biomarker. Additionally, the dynamic context of the regulatory environment may lead to reappraisal of the contexts for which a biomarker has been evaluated. For example, some regulatory environments may, despite attempts to minimize subjectivity, exhibit less caution when evaluating some contexts in which a given biomarker can be used. Thus, given the many demands and time constraints of the medical, scientific, and regulatory enterprises, the committee concludes that to incorporate and consider new research findings, biomarkers may be reevaluated within a reasonable time frame, such as every 4 years, for example. The committee does not intend such a time frame to dissuade more frequent reevaluation: Indeed, the rapidity of new knowledge available may dictate more immediate revisions in the contexts for which a biomarker may be used.

Rather, all biomarker evaluations should undergo reappraisal on at least such a time frame.

Each step needs to be reconsidered to the extent that research or context has changed since the previous evaluation. The reappraisal process need not consider the biomarker as though no previous evaluation had occurred. The monetary and opportunity costs of this kind of de novo evaluation would render such analyses prohibitive. Rather, the available data can be scrutinized in the context of what had been previously evaluated. By considering additional evidence, it is possible that the expert panel may alter its past findings by revoking recommendations for a previously accepted biomarker use, choosing not to recommend a biomarker for uses similar to those for which it was granted permission in the past, providing a more nuanced explanation as to how a biomarker should be used, or qualifying the biomarker for use in new contexts. Some of these scenarios are indicated in the case studies presented in Chapter 4. Nonetheless, it is essential that the utilization analysis be carried out by a panel of experts, as scientific and medical judgment is necessary to weigh the possible advantages and disadvantages of the proposed biomarker use.

SCIENTIFIC PROCESS HARMONIZATION

Recommendation 3:
The FDA should use the same degree of scientific rigor for evaluation of biomarkers across regulatory areas, whether they are proposed for use in the arenas of drugs, medical devices, biologics, or foods and dietary supplements. Congress may need to strengthen FDA authority to accomplish this goal.

Legislation and court decisions have created a regulatory environment in which different evidentiary and labeling requirements exist for drugs and biologics, devices, and foods and supplements. The committee has concluded that accurate and complete science is critical in all of these areas. While recognizing the differences between the different product categories, the committee emphasizes that none of these categories presents a situation so low in risk to consumers as to allow less rigorous scientific justification for claims. Box 3-6 summarizes the case study for a nutritionally relevant biomarker, blood levels of beta-carotene. This case study illustrates the need for collection of data for nutrition-related biomarkers.

To further illustrate the assertion that it is not safe to make assumptions about risks posed by products in a given category, consider the numbers of people exposed annually to several public health interventions that use food, compared to the numbers of people annually who

BOX 3-6
Blood Levels of β-Carotene

Studies have consistently shown that diets rich in fruit and vegetables are associated with a reduced risk of chronic diseases such as heart disease and cancer (Block et al. 1992; Peto et al., 1981). Although fruits and vegetables offer many nutrients, years of epidemiological studies suggested that blood levels of β-carotene were associated with lower incidence of cardiovascular disease and cancer (Hennekens et al., 1984; Manson et al., 1993; Willett et al., 1984). β-carotene is a carotenoid and antioxidant known to be a precursor of vitamin A.

To further corroborate the biomarker's biological plausibility, β-carotene's classification as an antioxidant provided a possible mechanism for a protective effect. Though there were no further animal studies or small-scale clinical trials performed, mounting pressures from multiple stakeholders, eager to prevent disease or improve the quality of life for persons at risk of chronic disease, quickly pushed the consideration of blood β-carotene levels as an effective chemopreventive biomarker and impelled large-scale intervention trials to test the possible benefits of increased intake of the nutrient itself were quickly initiated.

Before results from the three large β-carotene trials (the Physicians' Health Study) (Cook et al., 2000), the Beta Carotene and Retinal Efficacy Trial (CARET) (Omenn et al., 1996a, 1996b), and Alpha Tocopherol Beta Carotene Cancer Prevention Study (ATBC) (Albanes et al., 1996) had been confirmed, the belief in the "efficacy" of increased β-carotene intake became widespread based on the observational studies that demonstrated association, but not causality. This was based on the consistency, strength of association, dose–response gradient, and biological plausibility. Thus, the unfavorable and even deleterious results of the trials were surprising to physician, patient, research-scientist, and policy-maker proponents of β-carotene. These studies demonstrated that assumptions that β-carotene was a valid causal predictor of decreased lung cancer risk were in error and illustrate the public health value of proper preclinical research strategies and evaluative process before permitting claims. This matter is discussed in greater detail in the full case study found in Chapter 4.

take a few common drugs. About 184 million people drank fluoridated water in the United States in 2006, about 62 percent of the entire population (CDC, 2006). Commercially available cereal flours and related products, milk and other dairy products, and fruit juices and drinks can be fortified with vitamin D. Milk and cereals are most frequently fortified (Calvo et al., 2004). Dietary intakes of vitamin D in the United States range from about 4.2 to 5.4 μg per person per day (depending on age and sex), most of which is from fortified foods (Moore et al., 2005). Additionally, about 27 percent of the U.S. adult population took

a supplement containing vitamin D in 2002 (Kaufman et al., 2002). In 2002, it was reported that about 5.2 percent of the U.S. adult population was taking statins (Freemantle, 2002; Kaufman et al., 2002). The most commonly used medication, acetaminophen, was taken by about 23 percent of the U.S. adult population in a given week (Kaufman et al., 2002). Just over 1 percent of U.S. adults were taking fluoxetine hydrochloride (Prozac) (Kaufman et al., 2002). These are among the most used over-the-counter and prescription medications in the United States. From these and other similar data, it can be concluded that exposure to some public health interventions is much more prevalent than exposure to the most common medications.

Further, many individuals are not aware that public health interventions involving food are not risk-free. Chapter 4 shows the risks of beta-carotene supplementation. The example above highlights a topic discussed more fully in Chapter 2: in order to make informed decisions, individuals need access to complete information (see Chapter 2 section titled "Biomarkers and Communication Strategies at the FDA"). Nonetheless, the ability to interpret this information depends on numeracy, and individuals making complex decisions may benefit from professional advice (see Chapter 2 section titled "Numeracy"). However, professional advice is generally not sought for dietary decisions, for example. Further discussion of issues related to the use of biomarker data and its impact on subsequent health-related decisions was discussed in Chapter 2 (see section titled "Cognitive Biases and Impacts of Evidence Gaps").

Recommendation 3 is consistent with other recent efforts to improve the use of science at FDA and in European regulatory agencies. The renewed effort to strengthen the scientific base at FDA is discussed in Chapter 5 (see section titled "Tracking the Effects of Biomarker Use at the FDA"). Chapter 5 also goes into detail about the different requirements in different product areas. It discusses the use of regulatory authority and where better use may be needed. In order to implement this recommendation, the FDA will need to better implement some of its existing regulatory authority, and it may also need additional regulatory authority. Recommendation 3 is not meant to imply that an identical process be used across all of the centers. Instead, it means that rigorous, complete review of all available scientific evidence is necessary before regulatory decisions can be made. In the case of foods and supplements, for example, this may require Congress to enact legislation to allow the FDA to compel companies to gather and submit data relating to the safety and efficacy of proposed products and health claims, based on both the nutrients of interest alone and on the whole products within which they are contained.

Addressing Differences in Current Standards for Drugs, Biologics, Devices, Supplements, and Foods

Recommendation 4:
The FDA should take into account a nutrient's or food's source as well as any modifying effects of the food or supplement that serves as the delivery vehicle and the dietary patterns associated with consumption of the nutrient or food when reviewing health-related label claims and the safety of food and supplements. Congress may need to strengthen FDA authority to accomplish this goal.

Drugs, biologics, and devices are evaluated on the basis of the safety and efficacy of the entire product. The regulatory framework governing these products, foods, and supplements are explained in greater detail in Chapter 5. The committee concluded that for the utilization step of the biomarker evaluation framework, it is necessary to evaluate the biomarker's proposed use in terms of the entire product in all situations. In addition, the committee concluded that it is important to evaluate efficacy as well as safety of proposed biomarker uses. Legislation may be required to implement this recommendation.

Currently, the safety of new food substances is evaluated for the individual substances within the context of intended conditions of use, and not on a product-specific basis as is done for drugs. Validity of claims made with respect to foods and supplements can be made on the basis of single ingredients in foods. There are some restrictions on the amount of fat, saturated fat, cholesterol, and sodium that foods bearing health claims can contain, and also on the need for a minimum amount of vitamin A, vitamin C, calcium, protein, fiber, or iron for foods bearing claims. Nonetheless, although review of proposed health claims takes into account the relationship of the specific substance that is the subject of the health claim to the health outcome of interest, it may not adequately consider the modifications of the substance's effect on the disease outcome by other bioactive components in that food or the diet. For this reason, it is important to include an analysis of the connection between the biomarker and other factors associated with conditions that can affect its efficacy and safety in the qualification process.

In addition to the modifying effects of other material components of a food or supplement on the effect of a health claim based on a single ingredient, it is also important to consider the modifying effects of a health claim on the overall healthfulness of the diet. More research in this area is needed.

CONCLUSION

This approach to biomarker evaluation extends beyond reviewing the scientific literature to determine biomarker acceptance. The recommended comprehensive evaluation framework is a process by which consensus may be reached about the qualification of a biomarker and considers context-independent and context-dependent qualifications, as well as analytical validation. The committee finds it important to make analytical validation a necessary component to biomarker validation; without high-quality research data, biomarkers cannot be effectively used. Furthermore, it is important to know whether a biomarker has prognostic value and whether the science underlying its role in disease is well understood. Determining that a biomarker has prognostic value and a well-defined scientific basis, however, is distinct from knowledge that modifying the biomarker will bring about clinical benefit or harm. Utilization, the process of making assessments of whether a proposed biomarker is fit for the purpose for which it is being proposed, is the third essential component of the biomarker evaluation process. The committee concludes that these three steps therefore warrant separation to ensure each receives its full consideration. For decisions involving regulatory bodies, the committee recommends that an expert panel conduct the evaluation reviews. Biomarker evaluations need to be continually updated to reflect the current state of the science.

Importantly, the committee has recommended that the scientific information used to inform policy decisions regarding biomarkers should be equally rigorous across proposed uses and product categories. Finally, in the special case of foods and supplements, accommodations are needed to ensure that the entire food or supplement is taken into account when evaluating biomarkers for nutrition-related uses.

REFERENCES

Albanes, D., O. P. Heinonen, P. R. Taylor, J. Virtamo, B. K. Edwards, M. Rautalahti, A. M. Hartman, J. Palmgren, L. S. Freedman, J. Haapakoski, M. J. Barrett, P. Pietinen, N. Malila, E. Tala, K. Liippo, E. R. Salomaa, J. A. Tangrea, L. Teppo, F. B. Askin, E. Taskinen, Y. Erozan, P. Greenwald, and J. K. Huttunen. 1996. Alpha-tocopherol and beta-carotene supplements and lung cancer incidence in the alpha-tocopherol, beta-carotene cancer prevention study: Effects of base-line characteristics and study compliance. *Journal of the National Cancer Institute* 88(21):1560–1570.

Apple, F. S., M. Panteghini, J. Ravkilde, J. Mair, A. H. B. Wu, J. Tate, F. Pagani, R. H. Christenson, and A. S. Jaffe. 2005. Quality specifications for B-type natriuretic peptide assays. *Clinical Chemistry* 51(3):486–493.

Apple, F. S., A. H. B. Wu, A. S. Jaffe, M. Panteghini, R. H. Christenson, NACB Committee Members, C. P. Cannon, G. Francis, R. L. Jesse, D. A. Morrow, L. K. Newby, A. B. Storrow, W. H. W. Tang, IFCC Comittee on Standardization of Markers of Cardiac Damage Members, F. Pagani, J. Tate, J. Ordonez-Llanos, and J. Mair. 2007. National Academy of Clinical Biochemistry and IFCC Committee for Standardization of Markers of Cardiac Damage Laboratory Medicine Practice Guidelines: Analytical issues for biomarkers of heart failure. *Circulation* 116(5):e95–e98.

Berneis, K. K., and R. M. Krauss. 2002. Metabolic origins and clinical significance of LDL heterogeneity. *Journal of Lipid Research* 43(9):1363–1379.

Biomarkers Definitions Working Group. 2001. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology and Therapeutics* 69(3):89–95.

Block, G., B. Patterson, and A. Subar. 1992. Fruit, vegetables, and cancer prevention: A review of the epidemiological evidence. *Nutrition and Cancer* 18(1):1-29.

Califf, R. M., K. S. Pieper, W. R. Harlan 3rd, and K. L. Lee. 1992. Serum cholesterol in patients undergoing cardiac catheterization for suspected coronary artery disease: Diagnostic and prognostic implications. *Annals of Epidemiology* 2(1-2):137–145.

Calvo, M. S., S. J. Whiting, and C. N. Barton. 2004. Vitamin D fortification in the United States and Canada: Current status and data needs. *The American Journal of Clinical Nutrition* 80(Suppl):1710S–1716S.

CDC (Centers for Disease Control and Prevention). 2006. *Water fluoridation statistics for 2006*. http://www.cdc.gov/Fluoridation/statistics/2006stats.htm (accessed October 5, 2009).

CDER (Center for Drug Evaluation and Research). 2001. Guidance for Industry on Bioanalytical Method Validation: Availability. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM070107.pdf (accessed December 29, 2009).

Cochrane Collaboration. 2009. *Cochrane handbook for systematic reviews of interventions version 5.0.2* [updated September 2009]. Higgins, J. P. T., and S. Green (editors). http://www.cochrane-handbook.org/ (accessed March 15, 2010).

Cook, N. R., I. M. Le, J. E. Manson, J. E. Buring, and C. H. Hennekens. 2000. Effects of beta-carotene supplementation on cancer incidence by baseline characteristics in the Physicians' Health Study (United States). *Cancer Causes and Control* 11(7):617–626.

Couzin, J. 2008. Clinical trials and tribulations. Cholesterol veers off script. *Science* 322(5899): 220–223.

Fleming, T. R. 2005. Surrogate endpoints and the FDA's accelerated approval process. *Health Affairs*24(1): 67-78.

Fraser, C. G. 2001. *Biological variation: From principles to practice*. Washington, DC: AACC Press.

Freemantle, N. 2002. Medicalisation, limits to medicine, or never enough money to go around? Spending on preventive treatments that help a few is unaffordable. *BMJ* 324(7342):1241–1242.

Goodsaid, F., and F. Frueh. 2007. Biomarker qualification pilot process at the U.S. Food and Drug Administration. *AAPS Journal* 9(1):E105–E108.

Gupta, S., and J. de Lemos. 2007. Use and misuse of cardiac troponins in clinical practice. *Progress in Cardiovascular Diseases* 50(2):151–165.

Hennekens, C. H., M. J. Stampfer, and W. Willett. 1984. Micronutrients and cancer chemoprevention. *Cancer Detection and Prevention* 7(3):147-158.

Hill, A. B. 1965. The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine* 58:295–300.

ICH (International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use). 1994. *ICH Guidelines: Text on validation of analytical procedures, Q2A.* Geneva, Switzerland: ICH.

IOM (Institute of Medicine). 2007. *The future of drug safety.* Washington, DC: The National Academies Press.

IOM. 2009. *Conflict of interest in medical research, education, and practice*. Washington, DC: The National Academies Press.

IOM. 2010a. Accelerating rare diseases research and orphan product development. http://www8.nationalacademies.org/cp/projectview.aspx?key=49099 (accessed March 15, 2010).

IOM. 2010b. Standards for systematic reviews of clinical effectiveness research. http://www8.nationalacademies.org/cp/projectview.aspx?key=49124 (accessed March 15, 2010).

Kaufman, D. W., J. P. Kelly, L. Rosenberg, T. E. Anderson, and A. A. Mitchell. 2002. Recent patterns of medication use in the ambulatory adult population of the United States: The Slone Survey. *Journal of the American Medical Association* 287(3):337–344.

Koulman, A., G. Lane, S. Harrison, and D. Volmer. 2009. From differentiating metabolites to biomarkers. *Analytical and Bioanalytical Chemistry* 394(3):663–670.

Kristiansen, J. 2001. Description of a generally applicable model for the evaluation of uncertainty of measurement in clinical chemistry. *Clinical Chemistry and Laboratory Medicine* 39(10):920–931.

Lathia, C. D., D. Amakye, W. Dai, C. Girman, S. Madani, J. Mayne, P. MacCarthy, P. Pertel, L. Seman, A. Stoch, P. Tarantino, C. Webster, S. Williams, and J. A. Wagner. 2009. The value, qualification, and regulatory use of surrogate end points in drug development. *Clinical Pharmacology and Therapeutics* 86(1):32–43.

Lee, J. W., R. S. Weiner, J. M. Salistad, R. R. Bowsher, D. W. Knuth, P. J. O'Brien, J. L. Fourcroy, R. Dixit, L. Pandite, R. G. Pietrusko, H. D. Soares, V. Quarmby, O. L. Vesterqvist, D. M. Potter, J. L. Witliff, H. A. Fritche, T. O'Leary, L. Perlee, S. Kadam, and J. A. Wagner. 2005. Method validation and measurement of biomarkers in nonclinical and clinical samples in drug development: A conference report. *Pharmaceutical Research* 22(4):2495–2499.

Lee, J. W., W. Devanarayan, Y. C. Barrett, R. Weiner, J. Allinson, S. Fountain, S. Keller, I. Weinryb, M. Green, L. Duan, J. A. Rogers, R. Millham, P. J. O'Brien, J. Salistad, M. Khan, C. Ray, and J. A. Wagner. 2006. Fit-for-purpose method development and validation for successful biomarker measurement. *Pharmaceutical Research* 23(2):312–328.

Libby, P., and P. Theroux. 2005. Pathophysiology of coronary artery disease. *Circulation* 111(25):3481–3488.

Manson, J. E., J. M. Gaziano, M. A. Jonas, and C. H. Hennekens. 1993. Antioxidants and cardiovascular disease: A review. *Journal of the American College of Nutrition* 12(4): 426–432.

Moore, C. E., M. M. Murphy, and M. F. Holick. 2005. Vitamin D intakes by children and adults in the United States differ among ethnic groups. *The Journal of Nutrition* 135(10): 2478–2485.

NACB Writing Group Members, A. H. B. Wu, A. S. Jaffe, F. S. Apple, R. L. Jesse, G. L. Francis, D. A. Morrow, L. K. Newby, J. Ravkilde, W. H. Wilson Tang, R. H. Christenson, NACB Committee Members, R. H. Christenson, F. S. Apple, C. P. Cannon, G. L. Francis, R. L. Jesse, D. A. Morrow, L. K. Newby, J. Ravkilde, A. B. Storrow, W. H. Wilson Tang, and A. H. B. Wu. 2007. National Academy of Clinical Biochemistry Laboratory Medicine practice guidelines: Use of cardiac troponin and B-type natriuretic peptide or N-terminal proB-type natriuretic peptide for etiologies other than acute coronary syndromes and heart failure. *Clinical Chemistry* 53(12):2086–2096.

Omenn, G. S., G. E. Goodman, M. D. Thornquist, J. Balmes, M. R. Cullen, A. Glass, J. P. Keogh, F. L. Meyskens Jr., B. Valanis, J. H. Williams Jr., S. Barnhart, M. G. Cherniack, C. A. Brodkin, and S. Hammar. 1996a. Risk factors for lung cancer and for intervention effects in CARET, the Beta-Carotene and Retinol Efficacy Trial. *Journal of the National Cancer Institute* 88(21):1550–1559.

Omenn, G. S., G. E. Goodman, M. D. Thornquist, J. Balmes, M. R. Cullen, A. Glass, J. P. Keogh, F. L. Meyskens Jr., B. Valanis, J. H. Williams Jr., S. Barnhart, and S. Hammar. 1996b. Effects of a combination of beta-carotene and vitamin A on lung cancer and cardiovascular disease. *New England Journal of Medicine* 334:1150–1155.

Palou, A., C. Pico, and J. Keijer. 2009. Integration of risk and benefit analysis—the window of benefit as a new tool? *Critical Reviews in Food Science and Nutrition* 49(7):670–680.

Peto, R., R. Doll, J. D. Buckley, and M. B. Sporn. 1981. Can dietary beta-carotene materially reduce human cancer rates? Nature 290(5803):201–208.

Rasnake, C. M., P. R. Trumbo, and T. M. Heinonen. 2007. Surrogate endpoints and emerging surrogate endpoints for risk reduction of cardiovascular disease. *Nutrition Reviews* 66(2):76–81.

Ricos, C., V. Alvarez, F. Cava, J. V. Garcia-Lario, A. Hernandez, C. V. Jimenez, J. Minchinela, C. Perich, and M. Simon. 1999. Current databases on biological variation: Pros, cons and progress. *Scandinavian Journal of Clinical and Laboratory Investigation* 59(7):491–500.

Rizzo, M., and K. Berneis. 2007. Erratum: Low-density lipoprotein size and cardiovascular risk assessment (*Q J Med* (2006) 99(1-14)). *QJM* 100(2):147.

Rosner, W., R. J. Auchus, R. Azziz, P. M. Sluss, and H. Raff. 2007. Utility, limitations, and pitfalls in measuring testosterone: An Endocrine Society position statement. *The Journal of Clinical Endocrinology and Metabolism* 92(2):405-413.

Sargent, D. 2004. Disease-free survival (DFS) vs. overall survival (OS) as a primary endpoint for adjuvant colon cancer studies. http://www.fda.gov/ohrms/dockets/ac/04/transcripts/4037T2.htm (accessed March 15, 2010).

Shah, V. P., K. K. Midha, S. Dighe, I. J. McGilveray, J. P. Skelly, A. Yacobi, T. Layloff, C. T. Viswanathan, C. Edgar Cook, R. D. McDowall, K. A. Pittman, S. Spector, K. S. Albert, S. Bolton, M. Dobrinska, W. Doub, M. Eichelbaum, J. W. A. Findlay, K. Gallicano, W. Garland, D. J. Hardy, J. D. Hulse, H. Thomas Karnes, R. Lange, W. D. Mason, G. McKay, E. Ormsby, J. Overpeck, H. D. Plattenberg, V. P. Sha, G. Shiu, D. Sitar, F. Sorgel, J. T. Stewart, and L. Yuh. 1992. Analytical methods validation: Bioavailability, bioequivalence and pharmacokinetic studies. Sponsored by the American Association of Pharmaceutical Chemists, U.S. Food and Drug Administration, Internationale Pharmaceutique, Health Protection Branch (Canada) and Association of Official Analytical Chemists. *International Journal of Pharmaceutics* 82(1-2):1–7.

Staessen, J. A., J. G. Wang, and L. Thijs. 2003. Cardiovascular prevention and blood pressure reduction: A quantitative overview updated until 1 March 2003. *Journal of Hypertension* 21(6):1055–1076.

Swanson, B. N. 2002. Delivery of high-quality biomarker assays. *Disease Markers* 18:47–56.

Tardif, J.-C., T. Heinonen, D. Orloff, and P. Libby. 2006. Vascular biomarkers and surrogates in cardiovascular disease. *Circulation* 113(25):2936–2942.

Wagner, J. A. 2002. Overview of biomarkers and surrogate endpoints in drug development. *Disease Markers* 18(2):41–46.

Wagner, J. A. 2008. Strategic approach to fit-for-purpose biomarkers in drug development. *Annual Review of Pharmacology and Toxicology* 48:631–651.

Wagner, J. A., S. A. Williams, and C. J. Webster. 2007. Biomarkers and surrogate end points for fit-for-purpose development and regulatory evaluation of new drugs. *Clinical Pharmacology and Therapeutics* 81(1):104–107.

Willett, W. C., B. F. Polk, B. A. Underwood, M. J. Stampfer, S. Pressel, B. Rosner, J. O. Taylor, K. Schneider, and C. G. Hames. 1984. Relation of serum vitamins A and E and carotenoids to the risk of cancer. *New England Journal of Medicine* 310(7):430–434.

Williams, S. A., D. E. Slavin, J. A. Wagner, and C. J. Webster. 2006. A cost-effectiveness approach to the qualification and acceptance of biomarkers. *Nature Reviews Drug Discovery* 5(11):897–902.

Writing Group Members, F. S. Apple, A. H. Wu, A. S. Jaffe, M. Panteghini, R. H. Christenson, NACB Committee Members, R. H. Christenson, F. S. Apple, C. P. Cannon, G. S. Frances, R. L. Jesse, D. A. Morrow, L. K. Newby, A. B. Storrow, W. H. Tang, A. H. Wu, IFCC Committee on Standardization of Markers of Cardiac Damage (C-SMCD) Members, F. S. Apple, C. P. Cannon, A. S. Jaffe, F. Pagani, J. Tate, J. Ordonez-Llanos, and J. Mair. 2008. National Academy of Clinical Biochemistry and IFCC Committee for Standardization of Markers of Cardiac Damage Laboratory Medicine practice guidelines: Analytical issues for biomarkers of heart failure. *Clinical Biochemistry* 41(4-5):222–226.

Wu, Alan H.B. 2010. Standardization of assays for clinically important enzymes that have biologic variation: what is all the fuss about? *Clinical Chemistry and Laboratory Medicine* 48(3)299–300.

4

Case Studies

INTRODUCTION

The committee undertook case studies to illustrate the use of the recommended biomarker evaluation framework. Five case studies are presented in this chapter, each highlighting one or more aspects of the framework. The first case study is tumor volume in cancer, which highlights the need for rigorous analytical validation. The second case study is C-reactive protein (CRP), which highlights that data are crucial to ascertaining whether a biomarker can be more than a prognostic factor. The third case study is troponin, which highlights the utility of biomarkers for which sufficient data for use of the biomarker as a surrogate endpoint do not exist. The fourth case study is on low-density lipoprotein (LDL) and high-density lipoprotein (HDL) cholesterol, which highlights how even biomarkers frequently used as surrogate endpoints need to be carefully evaluated prior to each use. Finally, the fifth case study is on beta-carotene, which contains lessons for each step of the qualification framework. In particular, beta-carotene highlights the importance of biomarkers in nutrition-related settings.

Table 4-1 gives a brief summary of the results of the case studies. As can be seen, biomarkers are useful for a variety of purposes. In order for a biomarker to be used as a surrogate endpoint, however, a strong understanding of the causal pathways of the disease process and of an intervention's intended and unintended effects are usually needed. Achieving such understanding is a daunting challenge, and the committee acknowledges that it is infrequent that this understanding is achieved. The case

TABLE 4-1 Brief Summary of the Results of the Case Studies

Biomarker	Analytical Validity	Qualification	Utilization: Possible Uses	Utilization: Surrogate Endpoint Use
Tumor Size	No	Numerous trials exist with inconsistent findings on tumor shrinkage and clinical benefit	Data needed to improve analytical validity of one or more test	Current data do not support use as a surrogate endpoint
CRP	High-sensitivity tests available	RCTs and observational studies available; limited data on CRP's biological role in disease progression	Risk prediction; potential expansion of statin treatment to specific populations	Current data does not support use as a surrogate endpoint
Troponin	Validated tests are available for many uses. More sensitive tests are also being developed	Extensive data for acute troponin, limited data on chronic troponin; data collection is ongoing	Safety uses	Current data does not support use as a surrogate endpoint

LDL	Validated tests are available for many uses. More accurate tests are also being developed	Extensive data on LDL, both RCTs and observational studies; repeated use of LDL as a surrogate endpoint	Risk prediction	Data supports use of LDL as a surrogate endpoint for some cardiovascular outcomes for statin drug interventions, but not for all cardiovascular outcomes or other cardiovascular interventions, foods, or supplements
HDL	Validated tests are available for many uses. More accurate tests are also being developed.	Limited data; the biological role of HDL not fully understood	Risk prediction	Current data does not support use as a surrogate endpoint
Beta-carotene	Validated measures of blood serum beta-carotene levels are available for many uses	Extensive RCTs and observational trials are available	Uses include a biomarker of intake of fruits and vegetables and an effective intervention to address vitamin A deficiency	Current data does not support use as a surrogate endpoint

NOTE: CRP = C-reactive protein; HDL = high-density lipoprotein cholesterol; LDL = low-density lipoprotein cholesterol; RCT = randomized controlled trial.

studies chosen are generally ones for which a great deal of data already exist, and in a number of the case studies, the biomarkers have been discussed for several decades. However, the case studies illustrate that even in these situations the lack of sufficient data for surrogate endpoint status for the biomarkers is evident. Readers of these case studies may wonder what lessons can be gained toward prospective evaluation of biomarkers. For a newly discovered biomarker, it is likely that very little data will be available for review in the analytical validation and qualification steps of the evaluation framework. In these situations, the lack of data should be noted. During the utilization step of the framework, then, needs for further data are identified. After these data are collected, the evaluation process can be revisited until the data available support the use for which the biomarker is proposed.

It should be emphasized that these case studies are illustrative. Complete, rigorous, systematic reviews of the evidence base were not conducted by the committee. Each case study first introduces general information about the biomarker itself. Analytical validation, qualification, and utilization analyses are then discussed. Finally, a summary of the lessons learned through each case is given.

Biomarker Discovery and Development

Although many candidate biomarkers have been reported, few have been sufficiently evaluated to justify their use in developing drugs or making treatment decisions. This slow pace has been attributed to the challenges posed by the discovery and development processes. The discovery process is dependent on the technologies available to interrogate complex biochemistry of health and disease, and identifying differences that can be detected consistently in diverse populations (IOM, 2007). Advances in the fields of genomics and proteomics have made it easier to interrogate hundreds or even thousands of potential biomarkers at once, leading to large datasets requiring sophisticated analyses to identify individual biomarkers of interest, or patterns of markers. A recent IOM committee determined that realizing the full potential of biomarker-based tools is dependent on progress in biomarker discovery (IOM, 2007). However, technologies to identify and quantify proteins and metabolites have lagged behind methods to assess nucleic acids because of the diverse biochemical characteristics of the protein and metabolic products of the human genome. Beyond technology platforms, the committee also discussed the need to develop new software packages, algorithms, and statistical and computational models capable of integrating data from multiple inputs, such as proteomic or genomic data from the same samples.

Drug and diagnostic industries, along with academic researchers,

are involved in biomarker discovery activities. In drug development, biomarkers may be used in target validation, or in demonstrating that a potential drug target plays a key role in the disease process; early compound screening, identifying compounds with the most promise for safety and efficacy; pharmacodynamic assays to assess drug activity and select schedule/dose; patient selection; and surrogacy (IOM, 2007). Because therapeutics are generally only effective in a subset of patients, drug and diagnostic industries may develop (or in some cases, codevelop therapeutics and diagnostics) assays to assess which subset of patients would most benefit from a therapeutic. However, once a drug is approved, there is less financial incentive to develop biomarkers to guide treatment decisions because it would likely restrict the number of patients taking the drug.

TUMOR SIZE AS BIOMARKER FOR CANCER CLINICAL ENDPOINTS

Biomarkers play several roles in patient care in the context of cancer, as discussed in the Institute of Medicine's (IOM's) *Cancer Biomarkers* report (IOM, 2007). In patients who do not have a cancer diagnosis, biomarkers can be used for risk stratification, prevention of carcinogenesis in precancerous tissues, and screening for early-stage tumors. Biomarkers aid in making a diagnosis of cancer, classifying a particular patient's disease, and determining disease prognosis. In the context of a particular treatment, biomarkers are used for treatment stratification (treatment decisions based on patient characteristics), risk management (regarding adverse effects of a therapy), monitoring effectiveness or side effects of a therapy, and post-treatment disease surveillance. One metric used as a biomarker in cancer care, in the absence of or in conjunction with molecular markers, is tumor size measured with anatomic imaging, most meaningfully expressed in terms of tumor volume (Lin et al., 2008; Van Beers and Vilgrain, 2008).

Tumor response rates, defined by a change in tumor bulk, were commonly used for making decisions regarding approval of anticancer drugs in the 1970s, but in the mid-1980s, the Food and Drug Administration (FDA) added a requirement that a clinical survival benefit or quality-of-life benefit should be demonstrated. Because long trials are usually needed to demonstrate significant survival benefit and the demand for new anticancer drugs is always urgent, in 1996 the FDA extended *Accelerated Approval* under subpart H of the New Drug Application for drugs that are effective against serious or life-threatening diseases as measured by surrogate endpoints to anticancer drugs (HHS, 1996).[1] This included

[1] See http://www.fda.gov/ForConsumers/ByAudience/ForPatientAdvocates/SpeedingAccesstoImportantNewTherapies/ucm128291.htm#accelerated.

surrogate endpoints such as tumor size as it is represented in composite endpoints such as progression-free survival and time to progression. Accelerated approval is granted with the understanding that confirmatory evidence gathered in postmarket trials will lead to traditional approval of the drug, and a lack of such evidence may result in its removal from the market by the FDA.

Lathia et al. (2009) recently noted that "between 1992 and 2004, 22 applications for 18 anticancer drug or biologic agents were granted accelerated approval in the United States. These approvals were generally granted on the basis of end points such as overall response rate, time to progression, and disease-free survival. Of the 22 applications that received accelerated approval before January 2004, 6 were converted to regular approvals (i.e., demonstrated an effect on survival/outcome) whereas the remaining 16 were not converted to regular approvals; all these agents remain on the market."

While the outcome measured in phase III cancer trials is often overall survival, surrogate endpoints play a large role in evaluation of new therapeutic agents in phase II clinical trials (Ratain et al., 1993; Sargent et al., 2009; Scher et al., 2008; Seibert et al., 2007). A primary endpoint commonly reported in phase II trials for cancer therapeutics is response rate, defined in its most primitive form as tumor shrinkage. Unfortunately, phase II results based on tumor shrinkage are not always predictive of outcomes in phase III trials. In the case of agents with low response rates in phase II that go on to show an increase in progression-free survival or overall survival in phase III trials, speculation has been that this result may be due to tumor stabilization rather than tumor shrinkage by these therapeutic agents. This would suggest that although tumor shrinkage is an important variable to monitor, the way response rates are measured in phase II trials is failing to capture all clinically meaningful changes that should be considered in the drug evaluation process (Dhani et al., 2009; Llovet et al., 2008; Stewart, 2008; Weber, 2009).

Tumor size is an inconsistently defined biomarker often used for determining efficacy of cancer therapeutics (Marcus et al., 2009). Validation, qualification, and utilization analyses are complicated by use of multiple imaging platforms (hardware), nonstandardized acquisition and analysis protocols (software), dissimilar contrast agents and targeted imaging agents across trials and institutions, and inconsistent methods for measuring, calculating, and reporting tumor size.

Tumor Size: Analytical Validation

Tumor size measurements reported include tumor diameter, volume, and mass, as measured using anatomic imaging modalities such as mag-

netic resonance imaging (MRI), computed tomography (CT), ultrasound (US), and mammography (Strassburg et al., 2008). Validating use of tumor size as a biomarker is difficult because it is measured and defined in different ways depending on the imaging modality, the type of tumor, and the institution (Tran et al., 2004). Tumor size is sometimes expressed as diameter of the tumor in one or two views. Such values can also be used to approximate tumor volume using a spherical, cuboidal, prolate spheroid, or oblate spheroid model. However, many solid tumors are of irregular shape, and their volume can be best approximated by measuring tumor diameter in three (if possible) orthogonal views and using an elliptoid model to estimate tumor volume. A growing body of literature is advocating for the use of elliptoid modeling of tumor volume as the most meaningful representation of tumor size in terms of its accurate reflection of changes in tumor bulk confirmed by other volumetric measurements, such as water displacement and its correlation to clinical endpoints. However, some widely used standardized response criteria, such as the Response Evaluation Criteria in Solid Tumors (RECIST), employ a sum of the longest dimension recorded of each tumor when attempting to quantify disease burden (Eisenhauer et al., 2009; Gehan and Tefft, 2000).

A newer and more accurate approach to estimating tumor volume involves using two-dimensional tumor contours on sequential imaging slices to calculate volume in three dimensions. This technique can be used with MRI, CT, and positron emission tomography–computed tomography (PET–CT) images. Tumors are outlined on each slice manually or with automatic model-based segmentation and compiled to estimate gross tumor volume. This technique, particularly with implementation of automatic model-based segmentation to reduce interobserver discordance, provides a platform for accurately measuring tumor volume in a way that is reproducible and can be standardized relatively easily (Galanis et al., 2006).

Tumor mass can also be approximated using an estimation of tumor volume. This may be a useful metric in a laboratory setting where such quantities can be confirmed using ex vivo measures, or when tumor size is measured with anatomic imaging and tumor density is measured with functional imaging, such as PET, and the two measures are combined to estimate tumor mass. In cases, however, where mass is extrapolated from volume using an estimate of density, this calculation may introduce another source of error.

To further complicate measurement of tumor size, tumor borders are often poorly demarcated in highly invasive cancers, resulting in ambiguity about diameter length and interobserver discordance. For treatment evaluation the most emphasis should be put on reproducibility and accuracy of serial measurements; in this case reproducibility includes stan-

dardizing data collection across institutions and trials so that meaningful comparisons can be made between populations of patients. The variability in imaging platforms and techniques makes it unlikely that step one of the qualification framework is fulfilled given the current lack of standardization in the field. Adherence to American College of Radiology Appropriateness Criteria regarding appropriate modalities for imaging various types of cancer at different points along disease progression and treatment is one effort that could decrease discrepancies in data collection across trials (Böhm-Vélez et al., 2000; Fishman et al., 2000; Javitt, 2007). Because standardizing the hardware used at individual institutions may be difficult, it may be more feasible to standardize imaging acquisition and analysis protocols within a multicenter trial, and certainly within institutions (Grossi et al., 2004). Finally, some have explored the use of Bayesian analysis techniques to improve the accuracy of conclusions drawn from tumor images and other clinical data (Vokurka et al., 2002; Yang et al., 2003).

Tumor Size: Qualification

Because the growth of local or metastatic cancer cells can lead to the death of the host, it is biologically plausible that shrinkage of the existing tumor or prevention of further growth could serve as indications of biological and clinical benefit. However, many hypotheses exist regarding how cancer causes death in an organism (Lichtenstein, 2005). Some cancers cause death because cancer cells, much like parasites, compete with native tissue for nutrients, so that the organism essentially starves. Tumors frequently interfere with physiologic processes through mass effect, such as compression of vessels and other luminal structures or intracranial compression of brain tissue, or through invasion of normal tissue, which can result in clinical disease and death of the organism. Paraneoplastic syndromes and immune response to neoplastic cells also play a role in the mortality and morbidity of many types of cancer.

Given the contributions of these and other factors, the biological plausibility of using tumor size as a surrogate endpoint for evaluating disease progression and therapeutic efficacy in cancer is not entirely obvious. Smaller tumors tend to grow faster, so major shrinkage of tumor mass does not necessarily translate to prolonged survival (Citron, 2004; Hudis, 2005). Data have shown that tumor size may not correlate with long-term clinical outcome in some cancers, such as in locally advanced breast cancer, where lack of nodal involvement is predictive of disease-free survival and overall survival rates, but tumor size does not affect these rates (Beenken et al., 2003; Berruti et al., 2008). Additionally, real clinical benefit is not always accompanied by measurable reduction in

tumor size, as is the case with cytostatic drugs or agents that reduce the density of cells within a tumor but leave the tumor volume unchanged (Young et al., 1999). Even in the case of treatment with conventional cytotoxic drugs, initial tumor shrinkage is nearly always followed by tumor cell repopulation (Kerbel, 2006).

In the case of many biomarkers or surrogate endpoints, a causal role for the biomarker in the disease pathway is established. LDL, for example, is hypothesized to have a causal role in the atherosclerotic disease process, and while this has not been conclusively proven, LDL is measured as a biomarker of atherosclerotic cardiovascular disease and targeted pharmaceutically. Clearly tumor size is a different brand of surrogate endpoint from most molecular biomarkers in that increasing tumor size is viewed as a result of disease progression, not a causative factor. The exception to this mode of thinking about tumor size is tumors that secrete biologically active factors that promote proliferation via autocrine or paracrine signaling; in this case tumor growth may beget tumor growth while adequate vascular supply exists to support it (Imamoto et al., 1991).

In many studies tumor size is used as an indicator of response rate and for determining time to progression and disease-free survival (Ohara et al., 2002; Ollivier et al., 2007; Pugnale et al., 2003). While the link between tumor size and clinical benefit is less firm than what is traditionally required for associating a biomarker with a particular clinical endpoint (Therasse et al., 2006), use of tumor size as a biomarker in cancer has been rationalized by the serious nature of the disease and a lack of more solidly linked prognostic indicators. It is important to emphasize, however, that this rationalization is not universally accepted (Fleming et al., 2009). As will be discussed in Chapter 5, the in situations where it is deemed reasonable to permit marketing of drugs before clinical outcome evidence is available, it is important that this data be collected and analyzed through postmarket studies. In cancers where tumors shrink predictably in response to efficacious cytotoxic therapy, serial tumor-size measurements can provide insight into whether a new therapeutic agent or technique warrants further study or whether a particular patient or patient population is likely to benefit from that therapy (Henson et al., 2005; Husband et al., 2004; Kamel and Bluemke, 2002; Karrison et al., 2007). For example, in the case of locally advanced breast carcinoma treated with cytotoxic agents, tumor volume calculated using measurements taken with US, mammography, or MRI have been demonstrated to be prognostic and can also aid in selecting an effective treatment regimen (Berruti et al., 2005; Buijs et al., 2007; Cheung et al., 2003; Dose Schwarz et al., 2005; Eng-Wong et al., 2008; Hylton, 2006; Noterdaeme et al., 2009). Similarly, tumor volume is a critical measurement for monitoring and

directing local control of non-small-cell lung cancer (NSCLC) with radiation therapy.

The response rate often reported in phase II trials is based on an incorrect premise that tumor size is analogous or proportional to the number of tumor cells, as described in RECIST (Desar et al., 2009; Park et al., 2003; Tuma, 2006). The Choi Criteria, which were originally developed to assess tumor progression in gastrointestinal stromal tumors (GISTs), incorporate tumor size and density (measured with contrast-enhanced CT) into a metric of tumor progression. The Choi Criteria are a more sensitive measure of responsiveness to a particular therapy and have been demonstrated to more accurately predict overall survival in GIST than reduction in tumor size (Benjamin et al., 2007; Choi, 2005; Choi et al., 2007; Hohenberger and Wardelmann, 2006; Sevinc and Turhal, 2008; Stacchiotti et al., 2009).

The Southwest Oncology Group developed new criteria for evaluation of response in NSCLC that define response to therapy as anything other than progression. Patients who demonstrate a decrease in tumor size or who have stable disease are considered nonprogressive, and in NSCLC this measure of "disease control rate" is more predictive of overall survival than tumor shrinkage. The North Central Cancer Treatment Group and National Surgical Adjuvant Breast and Bowel Project (NSABP) have similarly used nonprogression at a specific time point as a measure of response to therapy that is more predictive of overall survival than tumor shrinkage (Tuma, 2006).

Tumor Size: Utilization

Cancer is a complex collection of diseases, which makes it difficult to make generalizations about how a particular surrogate endpoint should be used in trials for all types of cancer. One caveat to using tumor shrinkage as a surrogate endpoint is that it may not represent clinical benefit in all situations. In the case of GIST, progression usually occurs within the original tumor boundaries. Treatment of these tumors with Gleevec (imatinib) results in decreased cell density within the tumor and prolonged patient survival, but rarely shrinks measurable diameters of existing tumors to a significant degree. In this example Gleevec is thought to have both cytotoxic and cytostatic effects, and GIST cells are replaced by myxoid degeneration following cell death, both reasons why tumor size as a surrogate endpoint correlates poorly with clinical endpoints (Benjamin et al., 2007; Choi, 2005).

Factors to consider for contextual analysis of tumor size as a surrogate endpoint include the following: (1) when in a patient's treatment this variable is considered, and (2) for what purposes (Cademartiri et al., 2008; Christensen, 2008). In some cancers, tumor size is a useful diagnostic

and prognostic biomarker. For some cancers, though, imaging tumor size does not play a significant role in prognosis at the time of diagnosis. In the context of locally advanced breast cancer, for example, nodal involvement has a greater role in prognostication than tumor size (Beenken et al., 2003). NSABP has established criteria using pathologic complete response, which is defined as no evidence of malignancy on histologic analysis, instead of tumor shrinkage measured with anatomic imaging, to predict long-term prognosis over the course of disease. Obviously pathologic complete response cannot be evaluated for all cancers in all sites at all points along the history of the disease, which is why imaging has such an enormous role in monitoring response to therapy. In some types of breast cancer, for example, monitoring tumor size with imaging is tremendously useful for gauging efficacy of a particular therapy (Berruti et al., 2005, 2008; Buijs et al., 2007; Cheung et al., 2003; Eng-Wong et al., 2008; Hylton, 2006; Nicoletto et al., 2008).

Tumor Size: Lessons Learned

Although tumor shrinkage does not positively correlate with clinical benefit in all situations, the patchy qualification of tumor shrinkage as a surrogate endpoint for cancer trials has been tolerated by regulatory agencies for several reasons. Cancer as a family of diseases continues to result in high mortality and morbidity. Truly novel and efficacious therapeutics are not emerging as rapidly as society demands. Conditional approvals based on tumor size are not always followed by full approvals, but when measured correctly and used in the appropriate context, perhaps in conjunction with other variables like tumor density, tumor size is a useful parameter for detecting clinical benefit (Jensen et al., 2008; Monteil et al., 2009; Specht et al., 2007). Even so, use of tumor size as a surrogate endpoint for regulatory approvals is decreasing and is being replaced by other, better qualified surrogate endpoints. These surrogates, including progression-free survival, also require postmarket studies to connect the interventions to beneficial changes in clinical outcomes.

Tumor size as a surrogate endpoint highlights the many analytical validation issues of imaging biomarkers. Validation standards for imaging biomarkers should vary depending on their intended use as surrogate endpoints; criteria should be more stringent for the purposes of drug registration than for earlier stages of drug development. Emerging molecular and functional imaging technologies will likely provide tools to address some of the deficiencies of anatomic imaging in cancer discussed here (Funaioli et al., 2007; Goldstein et al., 2005; Pantaleo et al., 2008a, 2008b; Schepkin et al., 2006; Ullrich et al., 2008; Wahl et al., 2009). Combined with functional imaging technologies like PET and targeted molecular

agents, and with dynamic imaging technologies like perfusion and diffusion MRI, anatomic imaging in the future may serve as a more reliable surrogate endpoint in clinical trials for cancer (Carrió, 2008; Jennings et al., 2008; Leimgruber et al., 2006; Noterdaeme et al., 2009; Sharma et al., 2009; Stadler and Ratain, 2000; Stegger et al., 2008).

C-REACTIVE PROTEIN

Although cardiovascular disease mortality has fallen over the past century, cardiovascular disease (CVD) remains the leading cause of death in the United States (Mensah and Brown, 2007). The aging of the population, decline in the case-fatality rate of cardiovascular disease, and a relatively stable incidence of cardiovascular disease has also translated to a higher prevalence of cardiovascular disease in the United States (Pearson, 2007). An estimated 80 million American adults have one or more types of CVD, with an estimated 38 million of these cases in individuals 60 years or older (Lloyd-Jones et al., 2009). According to one estimate by the American College of Cardiology, the prevalence of chronic heart conditions will grow approximately 16 percent a decade for the next three decades (Foot et al., 2000). The burden of high prevalence of disease—in terms of morbidity, lost productivity, monetary cost, and increased use of healthcare services—has prioritized the need for improved prevention, risk assessment, and treatment of heart disease.

Traditionally, the prevention of cardiovascular disease has occurred through lowering risk factors associated with the development of CVD (Krumholz and Lee, 2008). The concept of risk factors was formalized by the Framingham Heart Study (FHS), a cohort study initiated in 1948 to assess the development of CVD over a long period of time in individuals who had not yet developed overt symptoms of CVD or suffered a heart attack or stroke (FHS, 2009; Kannel et al., 1961). Studies such as the FHS were able to demonstrate that hyperlipidemia and high blood pressure precede the development of CVD, and they are also associated with a higher risk of disease development. Compelling epidemiological and clinical trial evidence has demonstrated that smoking, hyperlipidemia, high blood pressure, and diabetes mellitus are independent risk factors for CVD, and therefore are considered the "traditional" risk factors for the disease (HHS, 1990; MacMahon et al., 1990; Stamler et al., 1993; Verschuren et al., 1995). Greenland et al. (2003) found that exposure to at least one clinically elevated traditional risk factor ranged from 87 to 100 percent for fatal coronary heart disease (CHD) in three prospective cohort studies while Khot et al. (2003) observed a prevalence of traditional risk factors of 80 to 90 percent of among patients with CHD.

Some have suggested that traditional risk factors (e.g., smoking,

hyperlipidemia, high blood pressure, and diabetes mellitus) do not fully explain cardiovascular risk. In the United States, each year 800,000 individuals will have a myocardial infarction (MI) and 700,000 will experience stroke, yet nearly half of these events occur in individuals without evidence of overt hyperlipidemia (Thom et al., 2006). On a population level, plasma total cholesterol levels poorly discriminate risk for coronary heart disease: 35 percent of CHD occurs among individuals with below-average levels of total cholesterol (Castelli, 1996). Khot et al. (2003) found that around 50 percent of subjects have zero or only one traditional risk factor for cardiovascular disease. In addition, many individuals with multiple risk factors never develop cardiovascular disease; Greenland et al. found that exposure to one or more of the traditional risk factors was highly prevalent among individuals who did not develop clinical CHD (Greenland et al., 2003), suggesting that additional work is needed to identify new risk prediction strategies. New biomarkers for cardiovascular disease are sought to improve risk prediction, to identify potential therapeutic targets, and to provide a more complete understanding of the pathophysiology of disease. Despite some controversy over the utility of new biomarkers in risk prediction (Greenland et al., 2003; Wang et al., 2006; Welsh et al., 2008), new biomarkers continue to be sought in hopes that the disease burden of cardiovascular disease can be mitigated by better identifying and stratifying those individuals at risk and intervening with better therapeutic targets.

As the understanding of cardiovascular disease has evolved to include the impact of inflammation on the progression of disease, inflammatory biomarkers have received substantial attention. Inflammation is believed to contribute to different stages in the pathogenesis of coronary heart disease, including a role in the development and progression of atherosclerosis (reviewed by Casas et al., 2008; Packard and Libby, 2008; Ross, 1993, 1999). A popular hypothesis, supported by both laboratory and clinical data, suggests that LDL modified by oxidation or glycation facilitates an inflammatory response in the artery wall, activating biological cascades that contribute to atherosclerosis initiation, progression, and complication (Packard and Libby, 2008; Ross, 1999). The most extensively studied inflammatory biomarker at the population level is C-reactive protein, but many other inflammatory biomarkers have been identified, including fibrinogen, serum amyloid A, VCAM-1, tumor necrosis factor, interleukin (IL)-1, IL-6, IL-18, and lipoprotein-associated phospholipase A_2, among others (Figure 4-1).

CRP is an acute phase, non-specific, systemic marker of inflammation. In normal individuals, CRP levels are low, but the serum concentration of CRP can increase upward of 1,000-fold upon exposure to a strong acute stimulus, such as sepsis or acute myocardial infarction (AMI), and

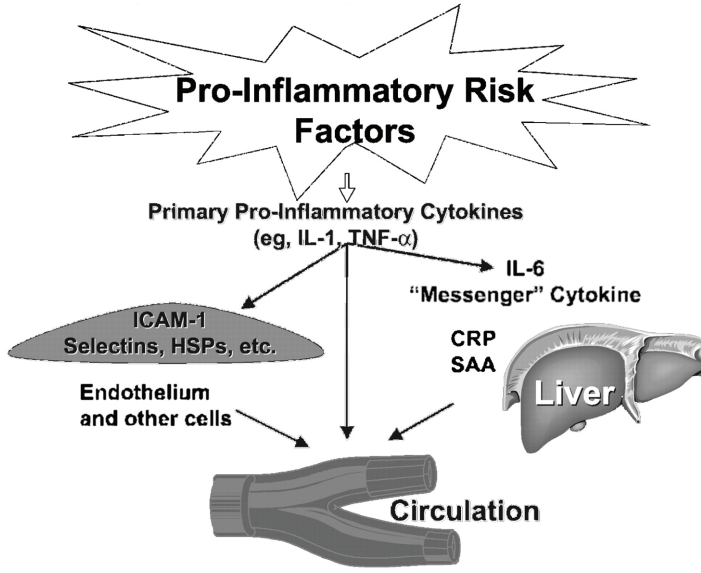

FIGURE 4-1 Inflammatory risk factors.
NOTE: CRP = C-reactive protein; HSP = heat-shock protein; ICAM-1 = intercellular adhesion molecule 1; IL-1 = interleukin 1; IL-6 = interleukin 6; SAA = serum amyloid A; TNF-α = tumor necrosis factor α.
SOURCES: Pearson et al. (2003). Reprinted with permission, Copyright 2003 by the American Heart Association. See also Libby and Ridker (1999).

can fall again when the stimulus is removed (Casas et al., 2008; Paffen and DeMaat, 2006). In the early 1990s, it was observed that individuals with active coronary syndromes and individuals with AMI who were tested prior to the acute-phase response to infarction were shown to have higher levels of CRP. The observation of the role of inflammation in cardiovascular disease and studies that revealed CRP predicted future coronary events began the current interest of CRP and cardiovascular disease (Pepys, 2005; Ridker et al., 1997).

CRP: Analytical Validation

CRP tests were first developed to measure acute phase responses of CRP, with detection limits around 2 to 10 mg/L. However, the newer assays were subsequently developed to measure CRP in the non-acute phase ranges, and are referred to as high sensitivity-CRP (hs-CRP) assays. These newer tests are the basis for measuring higher levels of CRP in the normal range associated with cardiovascular risk prediction, and com-

monly use CRP cutoffs of less than 1 mg/L, 1–3 mg/L, and greater than 3 mg/L to indicate low, average, and high cardiovascular risk.

CRP is easily measured via a number of standardized commercial hs-CRP assays, typically with detection limits of less than 0.3 mg/L and assay imprecision of less than 10 percent at low CRP concentrations (Roberts, 2004). In their 2003 scientific statement, the American Heart Association (AHA) and the Centers for Disease Control and Prevention (CDC) indicated that the hs-CRP assay was the best inflammatory assay candidate (Table 4-2) (Pearson et al., 2003). At the same time, the CDC published the first set of guidelines related to standardization of immunoassays for measurement of CRP (Kimberly et al., 2003). Sources of analytic variation for hs-CRP assays include laboratory methodology, reference material, precision, and calibration, among others (Ledue and Rifai, 2003). However, hs-CRP assay standardization efforts have continued (Kimberly et al., 2009), and the hs-CRP assay is currently considered standardized.

Specimen collection variables and physiologic characteristics are known to impact CRP measurement, but most research has indicated that CRP is a robust analyte that has negligible diurnal variation, does not depend on food intake, and has a long half-life (19 hours). Fresh, stored, and frozen plasma provide similar CRP measurement results (Ledue and Rifai, 2003). Physiologic characteristics, including race, ethnicity, age, sex, seasonality, biological variation, and lifestyle factors, have variable impact on CRP concentration. Some evidence suggests that men and women who are not receiving hormone replacement therapy have comparable CRP distributions (Ledue and Rifai, 2003), but other studies suggest that different gender subgroups have either lower or higher CRP concentrations. For instance, Japanese women may have slightly lower CRP concentrations (Yamada et al., 2001), while those of black females tend to be significantly higher (Albert et al., 2004). Likewise, research has also indicated that CRP concentrations vary by race and ethnicity, but there are limited data to evaluate the clinical relevance of these differences. Lifestyle factors that impact CRP levels include exercise, smoking, measures of adiposity, alcohol, anti-inflammatory drugs, and estrogen replacement therapy.

Despite these sources for variation in CRP concentration and measurement, CRP has proved to be a clinically useful measurement because it is an independent predictor of cardiovascular risk (Ridker, 2007), and there are widely available standardized, relatively low-cost hs-CRP assays that can be subjected to a number of collection variables, making CRP relatively easy for clinical use.

TABLE 4-2 Assays of Inflammatory Markers for Potential Clinical Use[a]

Analyte	Stability	Assay Availability	World Health Organization Standards Available?[b]	Interassay Precision
Soluble adhesion molecules (e.g., E-selectin, P-selectin, intracellular adhesion molecule-1, vascular cell adhesion molecule-1)	Unstable (unless frozen)	Limited	No	CV<15%
Cytokines (e.g., interleukin-1β, -6, -8, and -10 and tumor necrosis factor-α)	Unstable (unless frozen)	Few	Yes (Gaines Das and Poole, 1993)	CV<15%
Acute-phase reactants				
Fibrinogen	Unstable[c] (unless frozen)	Many	Yes[d] (Whitton et al., 2000)	CV<8%
SAA	Stable	One	Yes (Poole et al., 1998)	CV<9%
hs-CRP	Stable	Many	Yes (Whicher, 1998)	CV<10%
WBC count	Stable	Many	Yes	CV<3%

NOTES: CRP = C-reactive protein; CV = coefficient of variation; SAA = serum amyloid A; WBC = white blood cell.

[a] Courtesy of William Roberts, M.D., Ph.D.

[b] Information on specific standards is available at the following World Health Organization website: http://www.who.int.biologicals.

[c] In correctly anticoagulated blood, stable for at least 12 hours on ice or several hours at room temperature.

[d] World Health Organization standard available only for mass assay, not for functional assays most commonly in use.

SOURCE: Pearson et al. (2003). Reprinted with permission, Copyright 2003 by the American Heart Association.

CRP: Qualification

Inflammatory biomarkers for cardiovascular disease are newly emerging, and have less evidence supporting their use than traditional biomarkers, such as LDL cholesterol or blood pressure. CRP is the most extensively studied inflammatory biomarker, but evidence gaps prevent full understanding of the biomarker. Although research indicates that CRP is an independent predictor of cardiovascular risk, it is not known whether CRP plays a causal role in cardiovascular disease, which creates uncertainty about its use as a potential therapeutic target and surrogate endpoint.

Although inflammation is clearly involved in the development of atherosclerosis, researchers have not definitively ascertained whether CRP plays an active role in the disease process. The role of CRP in human physiology is not fully understood (Nilsson, 2005). Because no CRP deficiency or structural polymorphism has been reported, nor have therapeutic interventions specifically inhibited human CRP in vivo, the effects of absence, inhibition, and lack of function of CRP have yet to be tested (Casas et al., 2008). Moreover, it is uncertain whether CRP is a bystander in the cardiovascular disease process or whether it plays a causal role in the pathophysiology of disease.

CRP has been shown to have prothrombotic and proinflammatory properties (Bisoendial et al., 2005; Pasceri et al., 2000), including the perpetuation and amplification of inflammation and the immune response (reviewed by Calabro et al., 2009). There are some concerns that studies demonstrating proinflammatory and prothrombotic effects of CRP have been confounded by contamination (Packard and Libby, 2008; Pepys, 2005; Taylor et al., 2005; Van den Berg and Taylor, 2005). However, other studies have demonstrated that contaminant-free CRP preparations have direct atherogenic effects (Singh et al., 2005; Yaron et al., 2006).

CRP can bind selectively to LDL and very low-density lipoprotein (VLDL), suggesting that CRP could potentially be involved in atherosclerosis (de Beer et al., 1982; Pepys et al., 1985; Zhang et al., 1999), and experimental studies showed that CRP avidly binds to modified LDL, which accumulates in atherosclerotic plaques (Bhakdi et al., 1999). Based on these observations, models were developed to test the effects of CRP on cardiovascular outcomes in animal systems. Paul and colleagues (2004) demonstrated that human CRP transgenic apolipoprotein E-knockout mice had larger aortic atherosclerotic lesions than control mice. However, with this same model, Hirschfield and colleagues did not detect any proatherogenic or proinflammatory effects of transgenic expression of human CRP (2003, 2005). Pepys et al. (2006) developed a small-molecule inhibitor of CRP and demonstrated that administration of the inhibitor to rats undergoing AMI abrogated increase in infarct size and cardiac dys-

function produced by injection of human CRP. However, it is important to note the limitations of these preclinical models: rats and mice have extremely low levels of native CRP and "[i]ntroduction of human CRP into animals, wherein the protein is interacting with xenogenic molecules, cells, physiological and pathological processes, cannot be assumed to be a robust test for functions of human CRP in humans" (Casas et al., 2008).

The JUPITER trial (Justification for the Use of Statins in Prevention: An Intervention Trial Evaluating Rosuvastatin) found that in individuals with CRP concentrations of 2.0 mg per liter or higher but LDL levels of less than 130 mg per deciliter,[2] statin treatment significantly lowered the rate of first major cardiovascular events as compared to placebo (Ridker et al., 2008a). In a later analysis of the data, Ridker et al. (2009) found that JUPITER participants who achieved both LDL cholesterol level of less than 1.8 mmol/liter and CRP less than 1 mg/liter had a recorded 79 percent reduction in vascular event rates, while participants who achieved both LDL cholesterol level of less than 1.8 mmol/liter and CRP of less than 2 mg/liter had recorded a 62 percent reduction in vascular event rates. LDL cholesterol and CRP reductions were only weakly correlated with each other in this analysis. Although the JUPITER trial did not show that lowering CRP levels reduced cardiovascular risk, the trial does indicate that those patients with LDL levels of less than 130 mg per deciliter and CRP levels of greater than 2 mg per liter are at higher absolute risk and that rosuvastatin therapy resulted in a significant benefit in lowering cardiovascular events. Previous statin trials, including CARE (Cholesterol and Recurrent Events) and PRINCE (Pravastatin Inflammation/CRP Evaluation), also found that high CRP levels are significantly lowered with pravastatin therapy (Albert et al., 2001; Ridker et al., 1998). In the PRINCE study, no significant association was observed between baseline CRP and baseline LDL levels, end-of-study CRP and end-of-study LDL levels, or change in CRP and change in LDL levels over time; in linear regression analyses, only pravastatin therapy and baseline CRP levels were significant predictors of CRP reduction (Albert et al., 2001).

Some of the limitations of the JUPITER trial have been discussed in the literature, especially in relation to understanding the biological role of CRP. Hlatky (2008) noted that JUPITER trial entry criteria (apparently healthy men and women with LDL cholesterol of less than 130 mg per deciliter and CRP concentrations of 2.0 mg per liter or higher) provided only limited and indirect information about the biological role of CRP. The JUPITER trial did not compare subjects with CRP measurements of greater than 2.0 mg per liter with subjects having CRP measurements of

[2] Individuals with LDL levels less than 130 mg per deciliter are not considered in the drug treatment range by National Cholesterol Education Program guidelines (NCEP, 2001).

less than 2.0 mg per liter, nor did the trial compare the use of other markers of cardiovascular risk. Additionally, the trial did not evaluate whether individuals with CRP levels of less than 2.0 mg per liter would benefit from rosuvastatin treatment (Hlatky, 2008). However, the Air Force/Texas Coronary Atherosclerosis Prevention Study (AFCAPS/Tex-CAPS) found that lovastatin was ineffective among participants with a ratio of total to HDL cholesterol and a CRP level that were both lower than the median (Rikder et al., 2001). Ridker and colleagues (2009) acknowledge that it has not been determined the extent to which anti-inflammatory properties of statins affect clinical outcomes and whether these effects are independent of LDL cholesterol but suggest that it remains an intense area of research. Chan and colleagues (2009) note that the JUPITER trial provides "no results showing that C-reactive protein is an independent predictor of the relative or absolute benefit of therapy, since the treatment effects seen with rosuvastatin could have been mediated by reductions in low-density lipoprotein cholesterol."

Additional studies suggest CRP is not involved in the disease process, including Mendelian randomization studies. Genetic data provide an opportunity to assess the causality of biomarkers for disease. For a biomarker that has a causal role, the expected random distribution in a population of a polymorphism that determines high or low biomarker concentrations would be skewed in individuals depending on their disease status. Data from so-called "Mendelian randomization" studies are accumulating for several biomarkers such as CRP. Indeed, Zacho et al. (2008) found that although polymorphisms in the CRP gene are associated with markedly higher CRP levels, genetic polymorphisms in the CRP gene were not associated with an higher risk of ischemic vascular disease. Arriving at a similar conclusion, Elliot et al. (2009) conducted a larger genome-wide association study that found a lack of concordance between CRP genotypes and effect on coronary heart disease risk. While these studies do not discount the role of inflammation in cardiovascular disease pathogenesis, they cast doubt on the causal role of elevated CRP levels in cardiovascular disease (Elliott et al., 2009; Nordestgaard, 2009; Shah et al., 2009), prompting some to suggest that CRP-targeted drug development efforts should be abandoned (Kolata, 2009). However, others argue that Mendelian randomization findings may be limited by alternative explanations for epigenetic phenomena (Ogbuanu et al., 2009) and low predictive ability for genes related to biomarkers.

Although the biological role of CRP in cardiovascular disease remains uncertain, CRP has been shown to be an independent predictor of cardiovascular risk; at least 30 population cohorts (Shah and deLemos, 2009) have found that higher levels of CRP in the normal range are associated with higher risks for future coronary events, includ-

ing MI, ischemic stroke, peripheral vascular disease, and vascular death (Calabro et al., 2009; Musunuru et al., 2008; Ridker, 2007). In addition to individual studies, four meta-analyses have been conducted to assess CRP's independent predictive ability for future cardiovascular disease (Buckley et al., 2009; Danesh et al., 1998, 2000, 2004). Danesh et al. (2004) found that CRP is a relatively moderate predictor of coronary heart disease in comparison with traditional risk factors (such as higher LDL levels and cigarette smoking). The most recent meta-analysis (Buckley et al., 2009) concluded that strong evidence indicates CRP is independently associated with CHD events, and that moderate, consistent evidence suggests that adding CRP to risk prediction models improves risk stratification for those individuals initially at intermediate risk. The purpose of this meta-analysis was to assist the U.S. Preventive Services Task Force in determining whether CRP assessment should be incorporated into guidelines for cardiovascular risk assessment. Specifically, the authors focused on the potential benefit of adding CRP to models to improve risk stratification for those among intermediate risk because it would enable better tailoring of treatment decisions based on reclassification into either higher or lower risk categories. The meta-analysis found moderate evidence for adding CRP to prediction models for those at intermediate disease risk.

In spite of being an independent predictor of risk, some question the clinical utility CRP holds over the traditional risk factors (Folsom et al., 2006; Wang et al., 2006). The low incremental value that CRP and other new biomarkers have over traditional risk factors has been largely attributed to minimal impact on the area under the receiver operator curve, or c-statistic. However, others argue that this is an incorrect usage of the c-statistic, suggesting instead that reclassification into clinically relevant risk strata is a better way to assess prospectively the clinical impact of models (Cook, 2007, 2008; Cook et al., 2006). For example, adding CRP and family history to prediction models using traditional risk factors reclassifies 30 percent of individuals at intermediate risk into higher or lower levels of cardiovascular risk (Cook, 2008). The new risk strata were found to be better calibrated by comparing the predicted probabilities with the observed proportions within the reclassified categories.

CRP: Utilization

The third step of the committee's qualification framework is a contextual analysis of the available evidence about a biomarker with regard to the specific proposed use of the biomarker. As discussed in Chapter 2 of this report, biomarkers have many uses in both clinical care and drug development, including for risk stratification, prevention, screening,

diagnosis, prognosis, patient selection, and pharmacodynamics (see Table 2-1). Potential uses of CRP include risk prediction, prevention, drug development activities, and as a surrogate endpoint for drug or health claim approval. Each of these circumstances requires differing levels of evidence.

For clinical use in risk prediction, several factors would need to be considered. First, the prognostic value would be tantamount; elevations in CRP levels would need to be definitively linked with increased risk for cardiovascular events. Other factors that could play an important role in the qualification of a biomarker for risk prediction would include the strength of the biomarker risk prediction capabilities compared to other biomarkers that predict risk (especially those indicating inflammation). Use of CRP for risk prediction could also depend on the incremental clinical value of including the biomarker test within the other methods of assessing risk. For example, the addition of CRP and parental history significantly improves global cardiovascular risk prediction with the Reynolds Risk Score for Men (Cook and Ridker, 2009; Ridker et al., 2008b), but other evaluations question the utility of emerging biomarkers in risk prediction (Folsom et al., 2006; Wang et al., 2006). The complexity of cardiovascular disease, including the nature of competing risks within the general population, may favor the use of CRP in risk prediction because inflammation may be an important risk factor for different subpopulations, such as non-HDL cholesterol measurement is for those with familial hypercholesterolemia. A 2003 a joint scientific statement from AHA and CDC recommended against routine use of CRP in risk assessment for primary prevention of CHD, but supported its use in persons with a 10-year CHD risk of 10 to 20 percent (or those at intermediate risk of developing CHD disease), although the benefits of this strategy were unclear (Pearson et al., 2003). Other meta-analyses, such as Buckley et al. (2009), found moderate evidence for adding CRP to prediction models for those at intermediate disease risk.

A further context of use may be primary prevention, such as the expansion of statin treatment based on observed drops in CRP and cardiovascular event rates. As noted in Chapter 3, use of an intervention meant for primary prevention has an extremely low tolerance for risk. However, within this minimal tolerance for risk reduction of a very common serious chronic disease, more risk may be tolerated than for an intervention intended to prevent a less common or less serious disease. In the JUPITER trial, the study participants would not have received statin therapy under current treatment guidelines. Some argue that the results of this study suggested a potential role for expansion of statin therapy to 6.5 million adults with normal LDL and high CRP (Michos and Blumenthal, 2009). However, others argue that expanding treatment may require results from

more than one study, and take into account the cost effectiveness of the expansion as well as unintended risks of treatment. For example, based on the JUPITER data, those treated had significantly higher glycated hemoglobin levels and incidence of diabetes (Hlatky, 2008; Ridker et al., 2008).[3]

In drug development, there is a continuum of uses from discovery on one end to surrogate endpoints on the other end. A determination of the general category of use for which the biomarker is intended is necessary to consider the biomarker's utilization. For use in early drug development, lower levels of evidence are required. For example, qualification will most likely depend on a low level of biological plausibility—that interventions based on CRP levels make some mechanistic sense. At the other end of the spectrum is use as a surrogate endpoint. CRP is not currently utilized as a surrogate endpoint in drug or health claim approval. In spite of CRP's utility in cardiovascular risk prediction, its normal function and role in cardiovascular disease remains uncertain. The lack of understanding of CRP's biological role in human physiology has elicited controversy over assertions of CRP's causal role in cardiovascular disease. More research is needed to clarify determinants of CRP variation and utility in diverse populations. Although several interventions are known to lower CRP, it is unclear whether the effects of different interventions on CRP are consistently correlated with clinical outcomes. Based on these findings, in the utilization step of the evaluation framework, CRP would not currently qualify for the context of use as a surrogate endpoint, but it may be used in risk prediction in certain populations.

CRP: Lessons Learned

The CRP case study illustrates the importance of evidence accumulation to support different biomarker uses. Current research indicates that CRP may have utility in risk prediction, especially for those at intermediate risk of cardiovascular disease. For use in primary prevention, aside from the JUPITER trial, there is limited information to assess the benefit of intervening with statin therapy in individuals with high CRP levels, but normal LDL cholesterol levels. Although there are indications that reductions in CRP may contribute to clinical benefits, it is unclear whether CRP participates causally in the disease pathway. Multiple interventions

[3] A recent meta-analysis found that statin therapy is associated with a slightly increased risk of development of diabetes. The meta-analysis suggested a class effect and found that statin therapy was associated with a 9 percent increased risk for incident diabetes but suggested that the risk is low both in absolute terms and when compared with the reduction in coronary events (Cannon, 2010; Sattar et al., 2010).

are known to affect CRP levels, including statins, fibrates, exercise, and weight loss, but the benefits of these different interventions on clinical outcomes are still under evaluation. It is not clear whether these all confer valuable benefits on clinical outcomes, and that these benefits are all similar. Likewise, CRP variation in diverse populations and its predictive capacity in diverse racial and ethnic populations require further research. Incomplete understanding of CRP's normal function or its role in the disease process prevents the use of CRP as a surrogate endpoint at this time. Therefore, while CRP may be a useful biomarker for risk prediction, more evidence needs to be accumulated to establish further uses of CRP, both within clinical practice and regulatory decision making.

TROPONIN

Acute myocardial infarction is diagnosed through use of biomarkers, and cardiac troponin (cTn) is the biomarker that is best able to fulfill this task. Troponin is a protein involved in the function of cardiac and skeletal muscle function. Cardiac and skeletal troponins are proteins with three subunits; cardiac troponin contains cardiac troponin C (cTnC), cardiac troponin I (cTnI), and cardiac troponin T (cTnT). Several of the cTn subunits found in cardiac tissue are easily differentiated from the skeletal forms of the protein. Of the three subunits, cTnC is isomorphic with its skeletal counterpart, and so it is not used in cTn assays. Both cTnI and cTnT are distinguished through cTn-specific amino acid sequences near the N-terminus in the case of cTnI and both the N- and C-termini of cTnT. cTn assays utilize these unique characteristics through use of sequence-specific antibodies (Babuin and Jaffe, 2005). Though cTnT and cTnI, both subunits of cardiac troponin, are quantitatively different, from the clinical perspective, they are equivalent with the exception of renal failure.

The use of cTn for the assessment of myocardial infarction in suspected acute coronary syndrome (ACS) patients is ubiquitous and guideline driven. cTn is released only from heart tissue, and therefore elevations of cTn indicate recent myocardial damage. Nevertheless, elevated levels of cTn do not imply a cause of myocardial injury nor are the levels automatically suggestive of an acute coronary event (Wu and Jaffe, 2008). In addition to patients with ACS and MI, there are clinically important groups for whom measurements of cTn can aid in diagnosis and management. Its use in chronic settings is newer, and relies on developing high-sensitivity assays. However, there are already examples of the use of this testing with contemporary assays to assess drug safety and toxicity; these examples have not only proven diagnostically important, but also have assisted in exploration of interventions to address some toxic drug effects (Cardinale et al., 2000, 2002, 2006; Sandri et al., 2003).

Troponin: Analytical Validation

Cardiac troponin is the preferred biomarker to diagnose myocardial infarction and is accepted as the standard biomarker for this use by the American College of Cardiology, European Society of Cardiology, World Heart Federation, and AHA (Thygesen et al., 2007). These entities have defined an elevated cTn as being above the 99th percentile of a healthy population (Morrow et al., 2007; Thygesen et al., 2007). This percentile range is currently determined for each assay (Apple et al., 2007; Morrow et al., 2007; Panteghini et al., 2001). These groups also mandate interpretation of a rising or falling pattern. Calculating this for most assays relies almost entirely on analytic variation (the accuracy and precision of the test used) because contemporary assays are not sensitive enough to detect smaller changes in troponin levels. While troponin levels are different in different patient populations based on a variety of biological factors such as age, posture, and more, biological variation is generally not considered when troponin levels are measured, due to inadequate sensitivity of the assays. In addition, the reference populations for these measurements are variable and determined from convenience samples composed of various ages, and include individuals without cardiovascular disease, but with high concentrations of cTn for other reasons (Apple, 2009). Such individuals often have an adverse prognosis over time, so it is difficult to distinguish between an age-related increase in values and a subtle comorbidity such as silent myocardial damage. The prognostic data argue for the latter (Daniels et al., 2008a; Zethelius et al., 2006).

Various assays, including improved assays with higher analytical sensitivity, are becoming available for the measurement of cTnI and cTnT. In general, cTnI or cTnT are captured by specific antibodies onto surfaces or particles, tagged with the same specific antibodies from free solution, and then tagged with fluorescently labeled antibodies or another detection molecule. The extent to which new assays can decrease non-specific binding, increase binding or detection efficiency, or increase the lifetime or stability of the reagents is a major determinant of improved quality of the assays. Different assays, however, are not standardized because each uses a different set of antibody configurations to detect cTn in blood (Apple et al., 2007). Thus, the different assays recognize different epitopes and thus may measure different fragments or modified forms of the biomarker. In addition, all assays are calibrated differently (Apple, 1999; Katrukha et al., 1998). Furthermore, false positives related to fibrin (Roberts et al., 1997), and heterophilic and crossreacting human antimouse antibodies (Kaplan and Levinson, 1999) are of particular concern (Jaffe et al., 2000), although such problems are relatively infrequent. A new generation of cTn assays that are more precise at low concentrations (Apple, 2009) are being developed. The comparability of these assays with older assays

is being discussed by the FDA with a focus on ensuring accuracy at the 99th percentile and minimizing false-negative and false-positive findings (Apple, 2009). Some have proposed an assay-to-assay scorecard comparison and evaluated various assays and determined their acceptance designation based on imprecision at the 99th percentile.

Although some research assays have a low degree of imprecision at the 99th percentile, some of them may not be able to manifest high precision at many of the lower, more normal, values. For example, consider Patient Z, a 50-year-old man with a history of hypertension, hyperlipidemia, and mild diabetes. He had a mildly positive stress test in the past, but is otherwise asymptomatic. He started on a new antidiabetic agent 5 years ago. At that time, his ultra-sensitive cTn level was 3 pg/ml. It was 4 pg/ml 4 years ago, 7 pg/ml last year, and is now 9 pg/ml. First, most contemporary assays cannot measure these low concentrations. For all of the ultra-high-sensitivity assays, with one exception, these values are within the 99th-percentile reference population. However, it is unclear whether the change in these values exceeds analytical imprecision and biological variability until the third value (7 pg/ml last year) for some assays. In addition, if Patient Z has developed a new comorbidity such as heart failure, such a change might be expected.

Latini and colleagues (2007) suggest that troponin values on the high end of the normal range, at least in association with disease, has adverse prognostic significance for heart failure patients (Latini et al., 2007). Although these preliminary data suggest these assays have enormous potential for detecting disease at a very early stage, much of the necessary data validating such an approach is not yet in the public domain. Moving forward with standardized assays required establishing common 99th-percentile values using a healthy reference population that could provide standardized material for all assays (Apple, 2009). Although assays to measure cTn have not yet shown complete analytical validation, this case study will presume that several of the assays have adequate sensitivity, specificity, precision, and reproducibility, and the biomarker will be advanced to qualification.

Troponin: Qualification

Although measurable levels of cTn are indicative of cardiac injury, they are not "synonymous with an ischemic mechanism of injury" (Jaffe et al., 2000). The Dallas Heart Study demonstrated that it is not normal or healthy for individuals to have detectable cTnT levels. cTnT elevation was associated with congestive heart failure, left ventricular hypertrophy, diabetes mellitus, and end-stage renal disease (Wallace et al., 2006), though the higher cTnT levels did not appear to mark acute events. Longitudinal

research demonstrates that cTnI concentrations increase with age, even in patients without CHD, indicating "silent myocardial damage" (Zethelius et al., 2006).

Clinical data from several different trials show higher risk of mortality in individuals with elevated cTn with contemporary assays. Thus, it is likely that higher sensitivity assays will assist researchers in describing this area more accurately. cTn levels provide a threshold for higher risk for patients who present with acute ischemia. For example, the Rancho Bernardo Study revealed that older individuals (70 years of age and older), both men and women, with elevated troponins (greater than 99th percentile using contemporary assays) had an higher risk of all-cause and cardiovascular death (Daniels et al., 2008b). Similarly, the Uppsala study revealed that elderly individuals (65 years of age and older) with elevated cTnI, as measured with a highly sensitive Beckman assay, had impaired cardiac performance and higher cardiovascular risk (Eggers et al., 2008). Indeed, even values below the 99th-percentile value for the whole group—but above the value that might have been used had only younger individuals been included—were at higher risk. The Prevention of Events with ACE inhibition (PEACE) trial has shown that there is a positive correlation between cTnT levels measured with a high-sensitivity assay not yet released in the United States and both cardiac death and heart failure (Omland et al., 2009). Many of these values were below the putative 99th-percentile value. Thus, the assays continue to evolve. This evolution leads to the conclusion that utilization analysis (step 1c of the biomarker evaluation framework) and as-needed and continuing reevaluation of analytical validation and qualification (steps 1a and 1b), upon which utilization decisions are based, are essential.

High levels of cTn (greater than 99th percentile of the reference range) are associated with other causes of cardiac injury (e.g., cardiac surgery, pulmonary embolism, congestive heart failure [Missov et al., 1997], ablation [Katritsis et al., 1998], and myocarditis [Lauer et al., 1997]), as well as with non-cardiac diseases (e.g., sepsis [Guest et al., 1995; Spies et al., 1998], preeclampsia [Fleming et al., 2000], end-stage renal disease [Needham et al., 2004], extensive burns [Chen et al., 2000], high-dose chemotherapy [Cardinale et al., 2006], and stroke [James et al., 2000]), even with contemporary assays. This list will increase as more sensitive and more precise assays evolve. Box 4-1 highlights conditions associated with elevated cTn in the absence of overt ischemic heart disease. For many but not all of these latter conditions, elevated levels of cTn can be linked to direct toxic effects. Thus, although cTn is a highly sensitive biomarker of cardiac injury, it is not specifically an MI biomarker.

Though it is apparent that elevated cTn is associated with a higher risk of mortality, there is limited evidence that decreasing troponin levels

BOX 4-1
Conditions Associated with High Cardiac Troponins

Critically ill patients, especially those with diabetes, respiratory failure, gastrointestinal bleeding, and sepsis
High-dose chemotherapy
Primary pulmonary hypertension
Pulmonary embolism
Renal failure
Subarachnoid hemorrhage
Scorpion envenoming
Drug toxicity (e.g., Adriamycin, 5-fluorouracil, Herceptin, snake venoms, carbon monoxide poisoning)
Hypothyroidism
Burns, especially if total surface burn area is >30%
Infiltrative diseases (e.g., amyloidosis, hemochromatosis, sarcoidosis, scleroderma)
Acute neurological diseases (e.g., cerebrovascular accident, subarachnoid bleeds)
Vital exhaustion
Sepsis and septic shock
Stroke
Ultra-endurance exercise
Postoperative noncardiac surgery patients

SOURCES: Jaffe (2001), adapted from Ammann et al. (2004) and Wu and Jaffe (2008). Reprinted with permission, Copyright 2001, Springer Science + Business Media.

through interventions improves mortality risk. As these clinical trials indicate, patients with even small elevations in cTn can be identified as having higher risk (Hamm and Braunwald, 2000; Heidenreich et al., 2001). Results from the TACTICS-TIMI 18 trial indicate that these patients derive clinical benefit from an early invasive strategy of coronary angiography and revascularization (Cannon et al., 2001; Morrow et al., 2001). Thus, continued collection of data in this area will be useful to elucidate whether this principle will apply to other areas.

As indicated in Box 4-1, elevated cTn with contemporary assays is associated with high-dose chemotherapy and may be a predictor of chronic cardiotoxicity. The development of cardiotoxicity in a cancer patient is a strong indicator for discontinuing chemotherapy (Pai and Nahata, 2000). Therefore, preventing cardiotoxicity in cancer patients is important for both cardiac outcomes and therapeutic opportunities (Cardinale et al.,

2006). Cancer patients with elevated cTnI who were treated with enalapril, an ACE inhibitor that is thought to inhibit the development of oxygen free-radicals, were found to have an observable, significant reduction in the development of cardiotoxicity (Cardinale et al., 2006). These trials were not conducted using the high-sensitivity assays, however, and so it is not known whether smaller detectable cTnI elevations would have the same predictive capability.

This discussion indicates that cTn has prognostic value, but that there are limited data available for determining whether interventions targeting troponin impact outcomes in a broad set of diseases.

Troponin: Utilization

In 2007, the National Academy of Clinical Biochemistry (NACB) formed a committee to recommend guidelines for using cTn in etiologies other than ACS and heart failure. The committee determined that cTnT and cTnI could be used to risk-stratify patients with end-stage renal disease. cTnT is more frequently elevated for reasons that are unclear at present because, in general, cTnT and cTnI are otherwise generally equivalent clinically. Additionally, in patients with end-stage renal disease, changes in cTn elevation may be indicative of adverse prognoses, including coronary heart disease and death (Apple et al., 2002; DeFilippi et al., 2003). However, no therapeutic interventions are known to reduce cardiovascular risk based solely on the results of cTn testing in patients with end-stage renal disease (NACB Writing Group Members et al., 2007). The benefits and risks of such interventions are not fully defined (Scirica and Morrow, 2004). Thus, renal function is an important covariate in any cTn analysis. The NACB committee also determined that a high level of evidence suggests the measurement of cTn can define risk among patients who are critically ill, including those with sepsis (NACB Writing Group Members et al., 2007).

At present high-sensitivity cTn has been applied as a safety biomarker. Few data support other roles. Safety biomarkers are biomarkers used in preclinical (animal testing, often) and early clinical testing of drug and device candidates for a number of common types of toxicities. A recent IOM workshop explored the current status of development of safety biomarkers, especially in the organ systems of the heart, liver, and kidneys (IOM, 2009).

There are valuable uses of cTn as a biomarker, including as a risk biomarker in phase I studies to indicate safety problems with tested drugs and to collect further information about the biomarker. In the case above, it is not totally clear when the values rise above the critical threshold for risk. It may be that a change in troponin level within the normal reference

range is sufficient, or it may be that a value above the 99th percentile is required. Furthermore, if an increase is important, one needs to determine analytical and biological variation to confirm that the changes exceed normal physiology and analytic error. Therefore, this analyte is not quite ready for widespread clinical application in chronic disease settings.

Troponin: Lessons Learned

The most appropriate use identified for high-sensitivity troponin assays is as a safety biomarker in early clinical trials, such as phase I trials. Further information is necessary to further validate the preanalytic and analytic characteristics of these new high-sensitivity troponin assays. Information on biological variation (i.e, the reference change value that includes both analytical and biological variation) is needed. This includes not only the biology, but also the precision of the measurements. Finally, as in many of the case studies, context for the use of this biomarker is key. Due to evolving scientific understanding and test capabilities, it is not expected that the utility of these assays will be static.

LDL[4] AND HDL AS BIOMARKERS FOR CARDIOVASCULAR RISK

Lipoprotein particles are complex structures composed of lipids (chiefly cholesterol, phospholipids, and triglycerides), and proteins (apolipoproteins and others). The understanding of lipoproteins' structure, function, and role in CVD is shaped by more than 80 years of research (Cohn et al., 1946; Gofman et al., 1954; Macheboeuf, 1929; McNamara et

[4] It is important to highlight an issue with the commonly used terminology for cholesterol. A key point to recognize is that all major estimators of coronary risk do not include LDL cholesterol but rather total cholesterol-to-HDL ratio. Also, LDL cholesterol is rarely measured, most often being estimated from other fractions. The problem arises from the conflation of risk factors with biomarkers for benefit of intervention (i.e., surrogate endpoints). As is described in Chapter 1, these are not the same. Inclusion of total cholesterol and HDL cholesterol in global CVD risk prediction models obviates the need to include LDL cholesterol. It is also true that change in LDL cholesterol is more appropriate as a biomarker for effects of intervention on CVD risk. While it has some limitations in this regard, it is not reasonable to use total cholesterol or total cholesterol/HDL ratios as biomarkers for intervention since cholesterol is too non-specific (in particular, it can be decreased by lowering HDL cholesterol, which may not be beneficial); and we have no evidence that reducing the total cholesterol to HDL ratio by increasing HDL is beneficial. One step toward dealing with this dilemma, although imperfect, is to move from LDL to non-HDL cholesterol (total cholesterol minus HDL) as both a risk marker and treatment target. This at least addresses the two limitations of LDL cholesterol noted above. Although the terminology is awkward and has not caught on clinically, non-HDL cholesterol has been adopted in a limited way by the National Cholesterol Education Program Adult Treatment Panel III (NCEP ATPIII), and is on the list of issues for further discussion in the Adult Treatment Panel IV (ATPIV).

al., 2006; Olson, 1998; Pederson, 1947). By the mid-20th century, scientists had defined and measured all lipoprotein classes (Gofman et al., 1954), and preparative ultracentrifugation allowed for the separation of major lipoproteins (Havel et al., 1955). This cumulative research facilitated the development of clinical lab assays for LDLs and HDLs that could be applied in large epidemiologic studies and in clinical practice (Burstein and Samaille, 1958).

Since then, studies have established the cardiovascular risk associated with elevated concentrations of LDL cholesterol (LDL-C), abnormal proportions of LDL and HDL particles (Gofman and Lindgren, 1950; Gofman et al., 1950a, 1950b), and the reduction in cardiovascular risk associated with high HDL cholesterol (HDL-C) concentration (Robins, 2001; Scanu, 1966). That led to therapeutic interventions to induce reduction in cardiovascular disease risk by lowering LDL (Lipid Research Clinics Program, 1984a, 1984b), particularly through the use of 3-hydroxy-3-methylglutaryl coenzyme A (HMG CoA) reductase inhibitors (statins) (4S Study Group, 1994; Collins et al., 2002; Downs et al., 1998; LaRosa et al., 2005; Ridker et al., 2008a).

The FDA has used drug-induced LDL-C lowering as a surrogate endpoint for improved patient outcomes in approving new chemical entities; approval does not necessarily require confirmation of clinical benefits via outcomes studies. Because the FDA's Center for Drug Evaluation and Research (CDER) considers LDL-C to be a qualified surrogate endpoint, if a drug is safe and effectively lowers LDL-C, the drug's approval can be based on these lipid-lowering effects combined with the absence of a safety signal in the absence of definitive clinical outcome data (Nissen, 2008). Although a low level of HDL-C may signal a higher coronary heart disease risk than a moderately high LDL-C level (Robins, 2001), HDL-C has not yet qualified as a surrogate endpoint for cardiovascular risk because there is no evidence that HDL-raising interventions can improve outcomes. The FDA's Center for Food Safety and Nutrition also considers LDL-C as a qualified surrogate endpoint, so authorized health claims for cardiovascular disease risk reduction can be approved for foods containing ingredients that have been shown to lower LDL-C, with few restrictions on what other ingredients the food may contain and in the absence of information about the effect of the food on the clinical outcomes of those who consume it.

LDL and HDL: Analytical Validation

Cardiovascular biomarkers need to have not only prognostic value, but they also need to be readily measurable via standardized high-quality assays that can be correctly interpreted (Zaninotto et al., 2007). The

National Cholesterol Education Program of the National Institutes of Health has put forth precision and accuracy standards for LDL-C testing (Bachorik and Ross, 1995). As a result, the laboratory measurement of LDL-C is a reliable and reproducible measure most commonly determined using the Friedewald formula, which estimates LDL-C from measurements of total cholesterol, triglycerides, and HDL-C. Because this calculation does not directly measure LDL-C, it has limitations; more recent homogeneous assay methods are capable of directly measuring LDL-C (Nauck et al., 2002). However, the levels obtained by direct measurement of LDL-C do not show improved performance over those calculated using the Friedewald formula (Miller et al., 2002; Schectman et al., 1996; Yu et al., 2000), and the lower values obtained by this measurement may result in misclassification of CVD risk using current risk categories (Mora et al., 2009b). The CDC's Cholesterol Reference Method Laboratory Network certifies manufacturers of clinical diagnostic products that measure total cholesterol, HDL-C, and LDL-C and has certified five assays for measuring LDL-C.

Lipoprotein measurements other than LDL-C may also be considered in assessment of the degree of atherosclerosis and the endpoint of CVD-related morbidity and mortality. For some populations, there is increasing evidence that LDL particle number (LDL-P), small LDL particle concentration, non-HDL cholesterol, and apolipoprotein B (apoB) may have stronger associations with CVD risk than LDL-C (Contois et al., 2009; Cromwell et al., 2007; El Harchaoui et al., 2007; Jiang et al., 2004; Sniderman, 2002) because measurement of LDL-C does not consistently reflect levels of these other measures. The number of apoB-containing lipoprotein particles has been frequently reported to be more strongly associated with CVD risk than LDL-C (Barter et al., 2006; Sniderman and Marcovina, 2006). Thus, apoB may be a legitimate candidate for a biomarker for CVD risk (Miremadi et al., 2002). Also, there is less analytic variability in measuring apoB than in measuring LDL-C. Some data suggest that apoB, as well as LDL-P (another measure of LDL particle number), may have an important role in judging CVD risk in patients with elevated triglycerides and reduced HDL-C (El Harchaoui et al., 2007); validation and confirmation in other studies is needed.

Particles of LDL are heterogeneous, and smaller, denser LDL particles are more strongly related to risk than large LDL (Berneis and Krauss, 2002; Krauss, 1995; Krauss and Burke, 1982; Mora et al., 2009a; Musunuru et al., 2009; St-Pierre et al., 2005; Williams et al., 2003). However, most studies have suggested that peak LDL size measurement alone does not add information beyond that obtained by measuring LDL-C, triglyceride levels, and HDL-C (Campos et al., 1995; Mykkanen et al., 1999; Sacks and Campos, 2003).

Although HDL-C quantifies the amount of cholesterol contained within HDL particles, this measure does not necessarily correlate with the number of HDL particles or the differences between particle types. HDL particles can be classified according to their size and density as well as their protein composition; some evidence suggests these subclassifications may coincide with the functional properties of HDL (Joy and Hegele, 2008a, 2008b). Additionally, apolipoprotein A1 (apoA1) holds promise for measuring the correlation between HDL and CVD risk (Knopp et al., 2008). Until more is known about HDL particles and their functionality, standardization of HDL measurements other than HDL-C may not be feasible.

LDL and HDL: Qualification

LDL is associated with the disease process for atherosclerotic CVD because cholesterol is directly involved in the pathological disease process. The disease process, though, is influenced by numerous other mediators, including chemokines, cytokines, growth factors, proteases, adhesion molecules, hemostasis regulators, and receptors (Lopes-Virella and Virella, 2003; Virella and Lopes-Virella, 2008). These factors may also influence plaque progression and instability (Tardif et al., 2006). Additionally, genetic diseases, such as familial hypercholesterolemia, support the concept that CHD is influenced by high levels of LDL-C (Ballantyne, 2002).

The biological complexity of HDL structure and function complicates the ability to fully understand its beneficial mechanisms. The complex macromolecule possesses, to name a few, antioxidant, anti-inflammatory, antithrombotic, and endovascular properties (Griffin et al., 1999; Nofer et al., 2002; Steinberg et al., 1989), thereby challenging the isolation of specific HDL effects. Reverse cholesterol transport (RCT) is considered to be the primary mechanism by which HDL protects against atherogenesis (Rader, 2006). A current working model of RCT involves eight critical steps and several organs (e.g., liver, intestine, kidneys), enzymes (e.g., cholesteryl ester transfer protein [CETP], hepatic lipase), and lipoproteins and lipids. Any of these molecules could potentially be targeted when aiming to increase HDL-C levels. However, the optimum target for intervention is unknown (Joy and Hegele, 2008a). Furthermore, Tangier disease, in which individuals have a relatively low CVD mortality regardless of markedly reduced levels of HDL, complicates our understanding of the links between HDL and CVD (Oram, 2001).

Several observational studies have addressed the associations among age, cholesterol levels, and mortality in an older population. The Framingham Study, for example, showed that in the average population over

age 70, no statistical relationship is apparent among age, cholesterol, and mortality (Kronmal et al., 1993; Schatz et al., 2001) and that cholesterol may, in fact, be associated with longevity in individuals 85 years and older (Weverling-Rijnsburger et al., 1997). Other studies, though, have shown a reduction in the coronary heart disease death and nonfatal MI among individuals aged 70–82 years treated with pravastatin compared to placebo (Shepherd et al., 2002).

Epidemiologic studies have consistently demonstrated that higher CVD risk is also observed with lower levels of HDL-C (Knopp et al., 2008; Robins, 2001). Observational evidence indicates that raising HDL may independently reduce CVD risk (Gordon and Rifkind, 1989). However, while HDL-C is correlated with CVD risk in large studies, HDL-C levels are not necessarily indicative for individual patients due to the lack of understanding of the molecule's cardioprotective properties (Briel et al., 2009; Joy and Hegele, 2008a).

Epidemiologic studies describe increasing CVD risk with increasing LDL-C even when other risk factors are present (Knopp et al., 2008). Circumstantial evidence shows that greater decreases in LDL-C translate into greater positive effects than smaller decreases (Cannon et al., 2004), further indicating the strength of association between LDL-C and CVD risk. This association is not, however, equal across all ages and genders (Aronow, 2006; Berra, 2000; Knopp et al., 2005).

CVD is a complex disease with multiple determinants; LDL and HDL particles are only two of the involved components. Nevertheless, a large body of evidence supports LDL-C, in particular, as a biomarker that is extremely good for predicting CVD risk. Several decades of research have established a significant correlation between LDL-C and CHD (Castelli et al., 1986; Goldstein et al., 1973; Watanabe et al., 1985). There are examples where lower LDL-C clearly and consistently indicates lower risk for CVD, such as in familial hypercholesterolemia (van Aalst-Cohen et al., 2004). Numerous clinical trials have shown that reducing LDL-C decreases the incidence of CHD-related clinical events (4S Study Group, 1994; Cannon et al., 2004; Lipid Research Clinics Program, 1984c). These trials indicate that interventions can reduce CVD risk and provide credence that LDL-C accurately captures clinical impacts (Tardif et al., 2006). Statin trials in particular support the use of LDL-C as a biomarker for CVD risk. The West of Scotland Coronary Prevention Study (WOSCOPS) is one of the seminal studies establishing that the administration of a statin (pravastin) reduced cardiovascular mortality and morbidity (Shepherd et al., 1995). Other clinical studies have confirmed the value of secondary prevention using statins in patients who have documented CHD or who have suffered an acute cardiovascular event (4S Study Group, 1994; Collins et al., 2002; Schwartz et al., 2001). Thus, FDA-approved statins can make

labeling claims for the reduction of cardiovascular risk when they have specifically proven, through outcome studies, to reduce cardiovascular risk in addition to reducing LDL-C.

However, there is a good deal of controversy within research and patient care communities regarding limitations on LDL-C's utility as a biomarker in CVD. Targeting only LDL-C to reduce CVD risk may, in some populations and for some patients, miss other, potentially more relevant attributes of the disease pathway (e.g., fibrates). Much of the research examining the relationship between LDL-C and CVD risk demonstrates correlation between LDL-C levels and patient outcomes (4S Study Group, 1994; Cannon et al., 2004; Lipid Research Clinics Program, 1984c). But because CVD is a multifactorial disease process, and because many of the drugs used to treat it have pleiotropic effects (Liao and Laufs, 2005), one biomarker is unlikely to ever be a perfect surrogate endpoint for use in CVD clinical trials. Studies exist in which LDL-C was robustly decreased yet cardiac events were not reduced: specifically, estrogen replacement does result in lower LDL-C levels (Herrington et al., 2000; Hulley et al., 1998). Nonetheless, Hulley et al. concluded in 1998 that estrogen replacement therapy did not confer a cardiovascular benefit in a randomized clinical trial. Herrington et al. also concluded in 2000 that estrogen or estrogen plus medroxyprogesterone acetate did not provide any cardiovascular benefit in women with coronary atherosclerosis.

Intervention studies involving HDL-C have not been consistent, and there is little consistent evidence that raising HDL confers predicted benefit. However, there is evidence that raising HDL independently in apoA1 transgenic mice reduces atherosclerosis (Plump et al., 1992), and there has been intense interest in the development and application of pharmaceutical agents designed to raise HDL-C (Canner et al., 1986; The Coronary Drug Project Research Group, 1975; Frick et al., 1987; Rubins et al., 1999).

In the latter half of the 1980s, scientists described individuals and families that lacked CETP, resulting in raised HDL levels (Inazu, 1990; Tall, 1993). In about 1990, the hypothesis that raising HDL had an antiatherogenic effect began to be discussed in the scientific literature. In 1992, experiments demonstrated that rodents lacking CETP were found to have elevated HDL and were resistant to atherosclerosis (Jiang et al., 1992; Tall, 1993). Soon after these developments, Pfizer began development of a cholesteryl ester transfer protein inhibitor, torcetrapib. The Investigation of Lipid Management to Understand its Impact in Atherosclerotic Events (ILLUMINATE) trial investigated whether torcetrapib would decrease cardiovascular disease risk. The trial was prematurely terminated because of a higher risk of death and cardiac events, despite evidence of increasing HDL-C and decreasing LDL-C and imaging studies of carotid and

coronary vessels showed no benefit. While a known off-target effect of torcetrapib is an increase in blood pressure, the study was unable to rule out adverse or non-beneficial effects related to CETP inhibition (Barter et al., 2007), raising questions about whether raising HDL is an effective strategy for preventing CVD. A recent meta-analysis concluded that increasing HDL levels alone does not reduce risk of coronary heart disease events or deaths (Briel et al., 2009).

This notable example also illustrates that lowering LDL-C does not always correlate with improved patient outcomes; that the manipulation of lipid-processing pathways to lower LDL-C may result in a multiplicity of effects; and that when new drugs are used in human subjects, they may display unanticipated mechanisms of action that result in off-target effects (Barter et al., 2007).

Despite the disappointment of torcetrapib, several pharmaceutical interventions associated with increasing HDL-C appear to reduce atherosclerosis, principally through the administration of niacin (Canner et al., 1986; The Coronary Drug Project Research Group, 1975; Nofer et al., 2002), fibrates (Frick et al., 1987; Rubins et al., 1999), and reconstituted HDL (Nissen et al., 2003; Tardif et al., 2007). The 2001 Veterans Affairs High-Density Lipoprotein Intervention Trial (VA-HIT), treatment with gemfibrozil (fibric acid), decreased CVD without a change in LDL-C, but with an increase in HDL-C and a decrease in triglycerides (Robins, 2001). Thus, there is need for ongoing research in HDL-C elevation for the prevention of CVD risk.

As previously indicated, the statin trials consistently demonstrate that lowering LDL-C can reduce CVD risk, while other LDL-lowering interventions have not always resulted in lower CVD risk. Nevertheless, exceptions to this pattern indicate that LDL-C is not the only contributor to the CVD pathway; there is no consistent relationship between the decrease in LDL and the magnitude of risk reduction with statins (Hayward et al., 2006). The VA-HIT trial, for instance, decreased CVD risk without lowering LDL-C (Robins, 2001) and, as pointed out above, the ILLUMINATE trial decreased LDL-C levels, but increased CVD risk (Barter et al., 2007). In the ENHANCE trial (Effect of Combination Ezetimibe and High-Dose Simvastatin versus Simvastatin Alone on the Atherosclerotic Process in Patients with Heterozygous Familial Hypercholesterolemia trial), the additional reductions in LDL brought about through use of ezetimibe did not result in fewer cardiovascular events in the study population (Krumholz and Lee, 2008). The relationship between LDL-C and CVD risk is not perfect in all circumstances and for all populations.

There are also areas of inconsistency with regard to HDL-C and CVD risk. Low HDL is a component of atherogenic dyslipidemia, which is characterized by higher triglycerides, remnant lipoproteins, and small LDL

particle size; the magnitude of the contribution of HDL to CVD risk independent of this dyslipidemia has not been firmly established (Musunuru et al., 2009; Walter, 2009).

LDL and HDL: Utilization

An often-cited article published nearly three decades ago listed 243 coronary risk factors (Hopkins and Williams, 1981). Since then, the complexity of this list has increased (Hoefner, 2008). Consequently, a single biomarker (e.g., LDL-C or HDL-C) may be insufficient to accurately predict CVD risk for all patients. For example, as previously discussed, age and triglyceride level may complicate the non-linear dynamic between LDL-C and CVD risk.

There is a continuum of risk for CVD for which there are markers other than LDL-C and HDL-C. For example, combined hyperlipidemia, in which patients have elevated LDL, hypertriglyceridemia, and low HDL, may have a higher risk for CVD, for instance, than patients with normal LDL-C, but high CRP. For these latter patients, measurement of LDL-C alone will not provide full insight into their potential disease pathways. The benefits of statins may be in part related to pleiotropic effects (e.g., anti-inflammatory effects, decrease in triglycerides) and not solely due to lowering LDL-C (Nissen, 2008; Ridker, 2007). Furthermore, because lipoprotein transport consists of a system involving VLDL, LDL, and HDL, any intervention may affect all these lipoproteins (Knopp et al., 2008). The multiple effects of any treatment must be considered, as must the patient's individual risk profile.

If raising HDL is beneficial, the benefit may depend on the method used: weight loss, exercise, niacin, and fibrates appear to be beneficial but not specific to raising HDL; estrogen is not beneficial (despite lowering LDL); and raising HDL by CETP inhibition (i.e., torceptrapib) is inconclusive because of off-target effects and the inability to assess impact on CVD. Although there is modest evidence for the benefit on atherosclerosis of infusing native ApoA1 or modified ApoA1, it has proven difficult to devise a therapeutic intervention that specifically raises HDL without decreasing triglycerides and confounding the metabolic profile.

There is variable utility in using lipid profiles for prognostication at different points in the CVD spectrum and in different patient populations. In normotensive patients lipid profiles have been shown to be linked to arterial pressure and vessel stiffness (Marques-Vidal et al., 1996). Some studies have shown that in patients with essential hypertension, LDL-C and HDL-C independently predict risk of cardiovascular events (Verdecchia et al., 2004). In patients with atherosclerosis, some evidence

supports the measurement and modification of triglyceride levels to reduce risk for CVD events (Durrington, 1998; Manninen et al., 1992).

In patients with established CHD, things are a little more complicated. In one study low serum cholesterol was associated with an improved outcome in patients with CHD, but with a worse outcome in patients without CHD (Califf et al., 1992; Sakatini et al., 2005). In other studies lipid levels in patients with severe CHD were not predictive of survival (Feenstra et al., 1998), but initiation of statin therapy was associated with improved outcome (Horne et al., 2000). A study of lipid profiles in patients with angina pectoris found that reduced HDL-C was more strongly associated with higher coronary risk than other lipid measurements (Bolibar et al., 2000).

In the case where myocardial infarction has occurred, results vary somewhat in terms of the predictive value of lipid profiles. Some of the evidence points to worse outcomes with elevated lipid profiles in the first 24 hours post-MI, and advocate for statin use as early as possible to improve outcomes (Gorecki et al., 2004). HDL-C was highlighted as more independently predictive than LDL-C or total cholesterol in one study that examined the prognostic value of these variables in predicting recurrent events during acute coronary syndrome (Correia et al., 2009). In a study examining the role of biomarkers in premature MI (meaning in young patients), lipid profiles were shown to have no prognostic value for outcomes, and greater emphasis was placed on other variables such as smoking, ejection fraction, and serum homocysteine levels (Pineda et al., 2009).

LDL and HDL: Lessons Learned

As one of the FDA-qualified surrogate endpoints for CVD, LDL concentration is often viewed as the benchmark biomarker.

The evidence supporting LDL as a biomarker rests almost entirely on the measurement of LDL cholesterol even though this substance is only one part of the lipid transport system. Both apolipoprotein B and the quantity and the composition of LDL particles themselves have potential to be more accurate measures of LDL for some populations (Berneis and Krauss, 2002; Tardif et al., 2006), showing that even for qualified biomarkers, developing standard measures is an ongoing process.

The strength of LDL as a surrogate endpoint is not absolute due to the heterogeneity of cardiovascular disease processes, the heterogeneity of LDL-lowering drug effects, and the heterogeneity of LDL particles themselves. Because cardiovascular disease is a multifactorial chronic disease, a single risk factor for the disease (e.g., LDL-C) cannot fully account for all the variability that leads to a particular outcome (Libby

and Theroux, 2005; Tardif et al., 2006). This report's case study of CRP suggests that inflammation, for example, may also affect the cardiovascular disease pathway. Furthermore, age, gender, and genetic factors have been shown to complicate these already complex disease dynamics; as a result, lowering LDL-C can never be considered a "perfect" indicator across all population groups. That said, there is high probability that lowering LDL for several interventions decreases risk of cardiovascular disease (4S Study Group, 1994; Cannon et al., 2004; Knopp et al., 2008; Lipid Research Clinics Program, 1984c), and LDL, although not perfect, is one of the best biomarkers for cardiovascular disease.

Interventions to address a multifactorial disease introduce potentially unforeseen effects, particularly when the causal disease pathways, the mechanisms of action of the intervention, and the characteristics of the biomarker itself are not fully understood. High-density lipoprotein does not qualify as a surrogate endpoint for cardiovascular disease risk because these characteristics, particularly the latter, introduce high levels of variability.

BETA-CAROTENE

Beta-carotene (β-carotene), a pigment-producing molecule in the skin of several fruits and vegetables, is a member of the plant carotenoid family. β-carotene from dietary and supplemental sources is partially converted to vitamin A. Two µg of supplemental all-*trans*-β-carotene or 12 µg of dietary all-trans-β-carotene have a bioequivalency to 1 µg of all-*trans*-retinol or 1 retinol activity equivalent unit (IOM, 2000a). Due to this pro-vitamin-A activity, β-carotene in large enough amounts can prevent vitamin A deficiency. Vitamin A is an essential nutrient with important roles in normal vision, gene expression, reproduction, embryonic development, growth, and immune function (IOM, 2000a).

Although many publications use the terms vitamin A and β-carotene interchangeably, only β-carotene—but not preformed vitamin A—has been associated with potential antioxidant activity (IOM, 2000b). However, an IOM committee charged with evaluating the purported antioxidant nutrients concluded that although β-carotene and other carotenoids display in vitro antioxidant activity, the evidence that they act as in vivo antioxidants in humans is controversial (IOM, 2000b). Due to possible differences in antioxidant properties of the various carotenoids, it is important to specify which carotenoid(s) is being discussed.

Beyond the body of research done to evaluate a role for β-carotene in cardiovascular health and reduction of cancer risk, β-carotene has also been studied in other disease contexts. A supplement containing a combination of zinc, vitamin C, vitamin E, and β-carotene had a small

beneficial effect in the secondary prevention of some patients with diagnosed age-related macular degeneration (a leading cause of blindness in elderly people) (Chong et al., 2007). However, a specific effect attributable to β-carotene apart from the other nutrients in this multinutrient supplement cannot be made. Moreover, it is not possible to determine if the mechanism of action for this multinutrient supplement was due to its antioxidant properties or to other biological effects. Although β-carotene has also been studied as a possible treatment for cataracts, osteoarthritis, Alzheimer's disease, and cystic fibrosis, the results are inconclusive (Gritz et al., 2006; Heliovaara et al., 1994; Kneckt et al., 1992; Wood et al., 2001).

To date, the IOM has not defined a specific Dietary Reference Intake (DRI) value for either an adequate or a safe intake of β-carotene for the general population of apparently healthy people (IOM, 2000a, 2000b). It did specify biological activity equivalents for the conversion of different β-carotene sources to vitamin A (as noted above). They concluded that the only adverse event associated with high or long-term doses of β-carotene from food sources is yellowing of the skin (IOM, 2000b). Therefore, researchers were surprised when supplemental sources of β-carotene used in several clinical trials were associated with higher risk of lung cancer in participants who were smokers or former smokers (IOM, 2000b). The latter trial results were particularly surprising given that numerous observational studies had shown that people who frequently consumed β-carotene-containing fruits and vegetables had a lower incidence of cancer and CVD.

The history of the use of β-carotene as an intervention for prevention of disease and as a biomarker of protection against disease is discussed in detail in the qualification step of the evaluation framework, below. This extensive experience with β-carotene provides for application of the proposed framework to a well-studied biomarker for several conditions, specifically cancer, CVD, and eye diseases. The steps in this process are presented in the next three sections and will be illustrated with the rationale and findings from these studies.

Beta-Carotene: Analytical Validation

The initial observations of the inverse association between the consumption of fruits and vegetables and several conditions were based on measures of dietary intake. The measurement methods included food frequency questionnaires and 24-hour recalls. The association remained robust despite the differences in measurement approach. The intake of carotenoids, including β-carotene, was also related to the presence of disease. The concentration of carotenoids was measured in blood, as it is easily accessible, and found to convey accurate information about the

consumption of fruits and vegetables. Extensive experience with dietary and blood assessments led to the conclusion that "blood concentrations of carotenoids are the best biologic markers for consumption of fruits and vegetables" (IOM, 2000b).

At the time that the hypotheses about β-carotene and chronic disease risk were most popular and were the basis for funding decisions on intervention trials, little or no attention was paid to the question of the reproducibility, accuracy, and precision of the measurement of serum β-carotene. Moreover, because it was assumed that β-carotene's effectiveness was likely due to its in vivo antioxidant properties, a major limitation was the absence of a reliable prognostic assay for in vivo assessment of antioxidant activity that has clinical predictive value. There is no experimental technique that isolates β-carotene and measures its in vivo antioxidant capabilities when bombarded by several oxidative products. It is commonly regarded to reflect time-dependent β-carotene intakes, and by some researchers, to indicate body stores of β-carotene.

Concurrent with the development of associations between β-carotene intake (and blood levels) and cancer, cardiovascular disease, and retinal degeneration, research was directed to addressing the putative biologic mechanism of antioxidant activity. These measures endeavored to assess the potential to counteract oxidative damage occurring within nucleic acid or macromolecules resulting from the presence of free oxygen radicals in cells. For instance, oxidized LDL was found to produce atherosclerotic lesions in animals, and β-carotene could modulate this effect (Steinberg, 1997). Despite these links to oxidative potential and its modification, assays that measure the ability of an intervention to alter oxidative potential in humans proved inconsistent and unreliable. Because the collective experience indicated that these measures of oxidative potential could not be used in population studies, β-carotene measurement became the accepted standard for assessing a component of antioxidant exposure. However, β-carotene function is not limited to its antioxidant properties. Equally noteworthy, other antioxidants from various dietary sources exist and modulate cell activity. These considerations suggest that the antioxidant hypothesis regarding disease risk is not tested necessarily by β-carotene studies. In effect, the associations between β-carotene intake, blood levels, and disease are not supported by measures of the putative biologic process.

Beta-Carotene: Qualification

Several large intervention trials were undertaken based on the inverse relationships described in large observational and in vitro studies of cancer, CVD, and eye diseases (IOM, 2000b). The in vitro and animal studies,

regardless of the quantitative method used, consistently provided a plausible biologic basis to explain the prevention of the atherosclerotic process by supporting the idea of antioxidants limiting free radicals pervasive in CVD as well as cancer and retinal damage. For example, observational evidence including population studies found inverse relationships between dietary β-carotene and coronary heart disease (Liu et al., 2001). These findings were concordant with in vitro studies of β-carotene as an antioxidant. Operating under one or more notions of β-carotene as a risk biomarker, a surrogate endpoint, and a beneficial intervention, several large-scale randomized clinical trials were funded and designed to evaluate whether increasing β-carotene intake might lower the risk of cancer, CVD, and eye disease.

Three primary prevention clinical trials (mentioned below), which used β-carotene and/or other agents to augment baseline β-carotene serum levels, refuted the commonly accepted view about β-carotene. Two of the studies, initiated in populations at high risk of lung cancer by virtue of cigarette smoking or asbestos exposure, found increasing serum β-carotene increased the risk of lung cancer morbidity and mortality, whereas the other did not assert an advantageous or adverse effect regarding use of β-carotene in the prevention of cancer and CVD. Secondary CVD prevention trials also did not exhibit results of a beneficial or adverse impact in using the carotenoid. However, studies monitoring the impact of β-carotene on preventing eye diseases have produced conflicting findings.

The Alpha Tocopherol and Beta Carotene Prevention (ATBC) Study, begun in 1985, examined the incidence of lung cancer in 29,133 Finnish smokers. Researchers administered 20 mg of β-carotene, 50 mg of alpha-tocopherol, and a combination of the two supplements to three respective cohorts. They followed each subject for 5 to 8 years, reported an "excessive cumulative incidence of lung cancer was observed after 18 months (for the 7,278 men taking only β-carotene and 7,287 men who took a combination of β-carotene and alpha-tocopherol) and progressively thereafter, resulting in an 18% difference in incidence by the end of the study (95% CI = 3 to 36 percent, P = 0.01) between the participants who received β-carotene and those who did not" (Heinonen and Albanes, 1994). The number of new lung cancer cases was 876 (Heinonen and Albanes, 1994). Death increased by 8 percent for those in the β-carotene treatment group compared to the control (Heinonen and Albanes, 1994).

Further support of β-carotene's role in promoting increased risk came from another National Cancer Institute–supported study, the Beta Carotene and Retinol Efficacy Trial (CARET). Organized in 1983 and randomized in 1985, this multicenter, double-blind, randomized controlled trial assessed whether 30 mg of β-carotene and 25,000 IU of retinol decreased

the likelihood of developing and dying from primary lung cancer or cardiovascular disease in 18,314 smokers, former smokers, and workers exposed to asbestos. The incidence rates for lung cancer mortality and CVD mortality were significantly higher in the active treatment cohort compared to the control. The relative risk for developing lung cancer in the active treatment group was 1.28 (95% CI = 1.04 to 1.57, P = 0.02) compared to the placebo-controlled subjects. The study confirmed 286 primary lung cancers. The relative risk for lung cancer–associated death in the active treatment group was 1.46 (95% CI = 1.07 to 2.00), and the relative risk for all-cause death was 1.17 (95% CI = 1.03 to 1.33). Of the 388 reported cases of lung cancer, 254 people died. With respect to CVD mortality, the relative risk was 1.26 (95% CI = 0.99 to 1.61). CARET was terminated prematurely based on its disconcerting results and the prior ATBC findings obtained during the study. Follow-up continued for an additional 5 years (Omenn et al., 1996).

The Physicians Health Study I, consisting of a randomized two-by-two factorial trial of β-carotene to prevent cancer and aspirin to prevent coronary heart disease, enlisted 22,071 male physicians, some of whom were smokers. Participants received either 50 mg of β-carotene, 325 mg of aspirin, or placebo on alternate days (Hennekens et al., 1996).The aspirin arm of the study was concluded first with evidence for prevention of coronary heart disease, and the β-carotene arm was continued for a total of 12 years. Researchers ascertained that there was neither benefit nor harm from β-carotene supplementation with respect to morbidity or mortality. The relative risk of lung cancer in men who took β-carotene compared to healthy men who consumed the placebo was 0.98 (95% CI = 0.91–1.06). Nevertheless, 1,273 men in the β-carotene group developed lung neoplasms. β-carotene did not provide significant benefit or harm on the number of myocardial infarctions (468 in the β-carotene group vs. 489 in the placebo group), strokes (367 vs. 382), deaths due to cardiovascular causes (338 vs. 313), all-important cardiovascular events (967 vs. 972), or deaths from all causes (979 vs. 968) (Hennekens et al., 1996).

Recently, two studies of female health professionals have reported similar findings. The Women's Antioxidant Cardiovascular Study employed a factorial design to test 50 mg of β-carotene administered every alternate day, 500 mg of vitamin C daily, and 600 IU of vitamin E daily in the secondary prevention of CVD. The trial found no benefit or harm with β-carotene on a primary combined cardiovascular endpoints (RR = 1.02, CI = 0.92–1.13) or individual CVD endpoints (Cook et al., 2007). There were also no beneficial or harmful effects for vitamin C or vitamin A. The Women's Health Study, a primary prevention study of 39,876 women 45 years and older, found no benefit or harm for either cancer or CVD with β-carotene supplementation (Lee et al., 1999).

Several of these findings complemented those summarized in the IOM report in 2000 indicating that β-carotene did not modulate the risk for CVD (IOM, 2000b). While β-carotene was highly promoted prior to these studies, the lack of efficacy led researchers, physicians, and policy makers to recognize β-carotene is not an effective intervention for CVD or cancer.

During this period, the inverse relationship between fruit and vegetable intake and eye conditions (development of cataract and retinal disease) attracted the interest of vision researchers. Survey data from the first National Health and Nutrition Examination Survey was consistent with an inverse association between antioxidants and retinal disease (Goldberg et al., 1988). Intake of vitamin A was negatively correlated with macular degeneration (Goldberg et al., 1988). Experimental studies had found that visible and ultraviolet light can damage the retina through production of superoxide radicals, and it was proposed that antioxidants might protect vision by modulating the effect of superoxide radicals (Feeney and Berman, 1976; Goldberg et al., 1988). The Physicians Health Study I provided an opportunity to test the potential effect of an antioxidant, β-carotene. This study of 22,071 U.S. physicians found, after 12 years of follow-up of this cohort for eye conditions, that there was no benefit or harm with β-carotene supplementation (relative risk for β-carotene = 0.96, 0.78–1.20) (Christen et al., 2007).

In 1990, the National Eye Institute initiated a placebo-controlled randomized trial of antioxidant vitamins (including β-carotene) or zinc or a combination of zinc and antioxidant vitamins to determine the effects on eye cataracts and acute macular degeneration (AMD) when compared to placebo (AREDS Research Group, 2001). This study, the Age-Related Eye Disease Study (AREDS), was reported in 2001 and found that the combination of antioxidants and zinc reduced the risk of AMD in those suffering from early stages of the disease (AREDS Research Group, 2001); the odds ratio was 0.72 (0.52–0.98). There was no effect on cataract formation and no significant adverse effects overall were reported. It was advised that the use of this treatment was contraindicated in smokers because of the prior studies of β-carotene and lung cancer in smokers. The combination of the three antioxidants and zinc has become the standard of care to slow progression of AMD. However, a follow-up trial (AREDS 2) is under way to test whether an antioxidant combination that does not contain β-carotene is effective.

To date, increasing β-carotene serum levels has no demonstrated value in predicting the development of cancer or CVD. Neither the subjects in control cohorts of the large clinical trials who had baseline levels of β-carotene nor those who had elevated β-carotene were necessarily protected from a neoplasm or a cardiac event. For example, 443 of the 14,573

men not given β-carotene in the ATBC study developed cancer (Albanes et al., 1996). In the same trial, nearly equal numbers of men given β-carotene and those not given the supplement died from a cardiovascular disease that was not ischemic heart disease, hemorrhagic stroke, or ischemic stroke (Albanes et al., 1996). To a certain extent, however, some argue that the clinical trials supported the observational studies. Upon examination of the cumulative incidence of lung cancer after 0 to 18 months of randomization in the CARET trial, the observed incidence between active treatment and placebo groups was nearly identical (Omenn et al., 1996). The results of the Physicians' Health Study (PHS) trial conferred similar information. Of interest, the PHS was able to confirm a positive benefit for aspirin 325 mg every other day on prevention of CVD for a cohort of the same size as the Feeney and Berman (1976) β-carotene cohort. Given the size of the study population and the statistical power, it is apparent that β-carotene does not have prognostic value.

Prior to the major randomized clinical trials, the available evidence was limited to animal and observational studies. Nearly all of these studies were consistent in showing an association between serum β-carotene levels and chronic disease risk. Different methodologies and repetition of studies with β-carotene found similar results for the chronic diseases.

Beta-Carotene: Utilization

Prematurely, β-carotene was considered a biomarker that researchers and physicians could utilize for prognostic and predictive means. Many in the scientific and lay communities viewed the inverse association between β-carotene and cancer and CVD as indicative of the likelihood of developing either chronic disease. Additionally, some considered β-carotene useful for prevention given that it appeared as though it could reduce the onset, development, and death from these diseases. It was evident in ATBC and CARET that β-carotene in sufficient quantity could also cause adverse effects under certain conditions. Unfortunately, results from the large clinical trials lagged behind public endorsement of β-carotene as a beneficial intervention to prevent cancer and CVD. Hence, the context and application were incorrectly and prematurely assigned.

Nevertheless, β-carotene has been successfully used as a biomarker of intake of fruits and vegetables and an effective intervention to address vitamin A deficiency. Data supporting uses beyond these are still lacking.

Beta-Carotene: Lessons Learned

Regulatory agencies and the scientific community can draw several lessons from the application of β-carotene in the committee's framework.

First, appropriate study of the biomarker for qualification is essential. Conclusions about the benefits of β-carotene were made before phase III trials were concluded. Emphasis on thorough scientific research at the preclinical stage could be a cost-effective approach to preventing mismanagement of time, energy, and finances. The qualification component of the biomarker evaluation framework emphasizes that the scientific basis for a biomarker is robust, accurate, and consistent, too. Consideration of multiple pathways by which β-carotene (and other biomarkers) exerts influence on a disease state should be examined as part of the qualification step of the framework. Chronic diseases usually involve more than one pathway to trigger the condition, which suggests that researchers should investigate how a biomarker might be implicated in several pathways to better elucidate its role in chronic disease pathology. Also, designating the appropriate study population is crucial to correctly determining the context-specific effects of a biomarker.

In sum, β-carotene is a good biomarker of fruit and vegetable consumption, but not a good surrogate endpoint for interventions aimed at preventing chronic diseases.

REFERENCES

4S Study Group. 1994. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: The Scandinavian Simvastatin Survival Study (4S). *Lancet* 344(8934):1383–1389.

Albanes, D., O. P. Heinonen, P. R. Taylor, J. Virtamo, B. K. Edwards, M. Rautalahti, A. M. Hartman, J. Palmgren, L. S. Freedman, J. Haapakoski, M. J. Barrett, P. Pietinen, N. Malila, E. Tala, K. Liippo, E. R. Salomaa, J. A. Tangrea, L. Teppo, F. B. Askin, E. Taskinen, Y. Erozan, P. Greenwald, and J. K. Huttunen. 1996. Alpha-tocopherol and beta-carotene supplements and lung cancer incidence in the alpha-tocopherol, beta-carotene cancer prevention study: Effects of base-line characteristics and study compliance. *Journal of the National Cancer Institute* 88(21):1560–1570.

Albert, M. A., E. Danielson, N. Rifai, and P. M. Ridker. 2001. Effect of statin therapy on C-reactive protein levels: The Pravastatin Inflammation/CRP Evaluation (PRINCE), a randomized controlled trial and cohort study. *Journal of the American Medical Association* 286(1):64–70.

Albert, M. A., R. J. Glynn, J. E. Buring, and P. M. Ridker. 2004. C-reactive protein levels among women of various ethnic groups living in the United States (from the Women's Health Study). *American Journal of Cardiology* 93(10):1238–1242.

Ammann, P., M. Pfisterer, T. Fehr, and H. Rickli. 2004. Raised cardiac troponins: Causes extend beyond acute coronary syndromes. *British Medical Journal* 328(7447):1028–1029.

Apple, F. S. 1999. Clinical and analytical standardization issues confronting cardiac troponin I. *Clinical Chemistry* 45(1):18–20.

Apple, F. S. 2009. A new season for cardiac troponin assays: It's time to keep a scorecard. *Clinical Chemistry* 55(7):1303–1306.

Apple, F. S., M. M. Murakami, L. A. Pearce, and C. A. Herzog. 2002. Predictive value of cardiac troponin I and T for subsequent death in end-stage renal disease. *Circulation* 106(23):2941–2945.

Apple, F. S., A. H. B. Wu, A. S. Jaffe, M. Panteghini, R. H. Christenson, NACB Committee Members, C. P. Cannon, G. Francis, R. L. Jesse, D. A. Morrow, L. K. Newby, A. B. Storrow, W. H. W. Tang, IFCC Committee on Standardization of Markers of Cardiac Damage Members, F. Pagani, J. Tate, J. Ordonez-Llanos, and J. Mair. 2007. National Academy of Clinical Biochemistry and IFCC Committee for Standardization of Markers of Cardiac Damage Laboratory Medicine Practice Guidelines: Analytical issues for biomarkers of heart failure. *Circulation* 116(5):e95–e98.

AREDS Research Group (Age-Related Eye Disease Study Research Group). 2001. A randomized, placebo-controlled, clinical trial of high-dose supplementation with vitamins C and E, beta carotene, and zinc for age-related macular degeneration and vision loss: AREDS report no. 8. *Archives of Ophthalmology* 119(10):1417–1436.

Aronow, W. S. 2006. Management of hyperlipidemia with statins in the older patient. *Clinical Interventions in Aging* 1(4):433–438.

Babuin, L. and A. S. Jaffe. 2005. Troponin: The biomarker of choice for the detection of cardiac injury. *Canadian Medical Association Journal* 173(10):1191–1202.

Bachorik, P. S., and J. W. Ross. 1995. National Cholesterol Education Program recommendations for measurement of low-density lipoprotein cholesterol: Executive summary. The National Cholesterol Education Program Working Group on Lipoprotein Measurement. *Clinical Chemistry* 41(10):1414–1420.

Ballantyne, C. M. 2002. Familial hypercholesterolaemia: Optimum treatment strategies. *International Journal of Clinical Practice* 130(Suppl):22–26.

Barter, P. J., C. M. Ballantyne, R. Carmena, M. Castro Cabezas, M. J. Chapman, P. Couture, J. de Graaf, P. N. Durrington, O. Raergeman, J. Frohlich, C. D. Furberg, C. Gagne, S. M. Haffner, S. E. Humphries, I. Jungner, R. M. Krauss, P. Kwiterovich, S. Marcovina, C. J. Packard, T. A. Pearson, K. S. Reddy, R. Rosenson, N. Sarrafzadegan, A. D. Sniderman, A. F. Stalenhoef, E. Stein, P. J. Talmud, A. M. Tonkin, G. Walldius, and K. M. Williams. 2006. Apo B versus cholesterol in estimating cardiovascular risk and in guiding therapy: Report of the thirty-person/ten-country panel. *Journal of Internal Medicine* 259(3):247–258.

Barter, P., M. Caulfield, M. Eriksson, S. M. Grundy, J. J. P. Kastelein, M. Komajda, J. Lopez-Sendon, L. Mosca, J.-C. Tardif, D. D. Waters, C. L. Shear, J. H. Revkin, K. A. Buhr, M. R. Fisher, A. R. Tall, and B. Brewer. 2007. Effects of torcetrapib in patients at high risk for coronary events. *New England Journal of Medicine* 357(21):2109–2122.

Beenken, S. W., M. M. Urist, Y. Zhang, R. Desmond, H. Krontiras, H. Medina, and K. I. Bland. 2003. Axillary lymph node status, but not tumor size, predicts locoregional recurrence and overall survival after mastectomy for breast cancer. *Annals of Surgery* 237(5):732–739.

Benjamin, R. S., H. Choi, H. A. Macapinlac, M. A. Burgess, S. R. Patel, L. L. Chen, D. A. Podoloff, and C. Charnsangavej. 2007. We should desist using RECIST, at least in GIST. *Journal of Clinical Oncology* 25(13):1760–1764.

Berneis, K. K., and R. M. Krauss. 2002. Metabolic origins and clinical significance of LDL heterogeneity. *Journal of Lipid Research* 43(9):1363–1379.

Berra, K. 2000. Women, coronary heart disease, and dyslipidemia: Does gender alter detection, evaluation, or therapy? *Journal of Cardiovascular Nursing* 14(2):59–78.

Berruti, A., D. Generali, M. P. Brizzi, M. Ardine, L. Dogliotti, and A. Bottini. 2005. Is pathologic complete response a valid surrogate parameter of treatment efficacy in HER2 positive breast cancer patients undergoing primary chemotherapy plus trastuz[u]mab? *Journal of Clinical Oncology* 23(31):8130–8132.

Berruti, A., M. P. Brizzi, D. Generali, M. Ardine, L. Dogliotti, P. Bruzzi, and A. Bottini. 2008. Presurgical systemic treatment of nonmetastatic breast cancer: Facts and open questions. *Oncologist* 13(11):1137–1148.

Bhakdi, S., J. Torzewski, M. Klouche, and M. Hemmes. 1999. Complement and atherogenesis. Binding of CRP to degraded, nonoxidized LDL enhances complement activation. *Arterioscler Thromb Vasc Biol* 19(10):2348–2354.

Bisoendial, R. J., J. J. P. Kastelein, J. H. M. Levels, J. J. Zwaginga, B. van den Bogaard, P. H. Reitsma, J. C. M. Meijers, D. Hartman, M. Levi, and E. S. G. Stores. 2005. Activation of inflammation and coagulation after infusion of C-reactive protein in humans. *Circulation Research* 96(7):714–716.

Böhm-Vélez, M., E. Mendelson, R. Bree, H. Finberg, E. K. Fishman, H. Hricak, F. Laing, D. Sartoris, A. Thurmond, and S. Goldstein. 2000. Ovarian cancer screening. American College of Radiology. ACR Appropriateness Criteria. *Radiology* 215(Suppl):861–871.

Bolibar, I., A. Von Eckardstein, G. Assmann, S. Thompson, ECAT Angina Pectoris Study Group, and European Concerted Action on Thrombosis and Disabilities. 2000. Short-term prognostic value of lipid measurements in patients with angina pectoris. The ECAT Angina Pectoris Study Group: European Concerted Action on Thrombosis and Disabilities. *Thrombosis and Haemostasis* 84(6):955–960.

Briel, M., I. Ferreira-Gonzalez, J. J. You, P. J. Karanicolas, E. A. Akl, P. Wu, B. Blechacz, D. Bassler, X. Wei, A. Sharman, I. Whitt, S. Alves da Silva, Z. Khalid, A. J. Nordmann, Q. Zhou, S. D. Walter, N. Vale, N. Bhatnagar, C. O'Regan, E. J. Mills, H. C. Bucher, V. M. Montori, and G. H. Guyatt. 2009. Association between change in high density lipoprotein cholesterol and cardiovascular disease morbidity and mortality: Systematic review and meta-regression analysis. *BMJ* 338:b231.

Buckley, B. M., R. Fu, M. Freeman, K. Rogers, and M. Helfand. 2009. C-reactive protein as a risk factor for coronary heart disease: A systematic review and meta-analyses for the U.S. Preventive Services Task Force. *Annals of Internal Medicine* 151(7):483–495.

Buijs, M., I. R. Kamel, J. A. Vossen, C. S. Georgiades, K. Hong, and J. F. Geschwind. 2007. Assessment of metastatic breast cancer response to chemoembolization with contrast agent enhanced and diffusion-weighted MR imaging. *Journal of Vascular and Interventional Radiology* 18(8):957–963.

Burstein, M., and J. Samaille. 1958. Determination of serum beta-lipoproteins after selective precipitation by heparin. *La Presse Médicale* 66:974–975.

Cademartiri, F., G. Luccichenti, E. Maffei, M. Fusaro, A. Palumbo, P. Soliani, M. Sianesi, M. Zompatori, G. Crisi, and G. R. Krestin. 2008. Imaging for oncologic staging and follow-up: Review of current methods and novel approaches. *Acta Bio-Medica* 79(2):85–91.

Calabro, P., E. Golia, and E. T. Yeh. 2009. CRP and the risk of atherosclerotic events. *Seminars in Immunopathology* 31(1):79–94.

Califf, R. M., K. S. Pieper, W. R. Harlan Jr., and K. L. Lee. 1992. Serum cholesterol in patients undergoing cardiac catheterization for suspected coronary artery disease: Diagnostic and prognostic implications. *Annals of Epidemiology* 2(1–2):137–145.

Campos, H., G. O. Roederer, S. Lussier-Cacan, J. Davignon, and R. M. Krauss. 1995. Predominance of large LDL and reduced HDL2 cholesterol in normolipidemic men with coronary artery disease. *Arteriosclerosis, Thrombosis, and Vascular Biology* 15(8):1043–1048.

Canner, P. L., K. G. Berge, N. K. Wenger, J. Stamler, L. Friedman, R. J. Prineas, and W. Friedewald. 1986. Fifteen year mortality in Coronary Drug Project patients: Long-term benefit with niacin. *Journal of the American College of Cardiology* 8(6):1245–1255.

Cannon, C. P. 2010. Balancing the benefits of statins versus a new risk—diabetes. *The Lancet* 375(9716):700–701.

Cannon, C. P., W. S. Weintraub, L. A. Demopoulos, R. Vicari, M. J. Frey, N. Lakkis, F. J. Neumann, D. H. Robertson, P. T. Delucca, P. M. Dibattiste, C. M. Gibson, and E. Braunwald. 2001. Comparison of early invasive and conservative strategies in patients with unstable coronary syndromes treated with the glycoprotein IIb/IIIa inhibitor tirofiban. *New England Journal of Medicine* 344(25):1879–1887.

Cannon, C. P., E. Braunwald, C. H. McCabe, D. J. Rader, J. L. Rouleau, R. Belder, S. V. Joyal, K. A. Hill, M. A. Pfeffer, and A. M. Skene. 2004. Intensive versus moderate lipid lowering with statins after acute coronary syndromes. *New England Journal of Medicine* 350(15):1495–1504.

Cardinale, D., M. T. Sandri, A. Martinoni, A. Tricca, M. Civelli, G. Lamantia, S. Cinieri, G. Martinelli, C. M. Cipolla, and C. Fiorentini. 2000. Left ventricular dysfunction predicted by early troponin I release after high-dose chemotherapy. *Journal of the American College of Cardiology* 36(2):517–522.

Cardinale, D., M. T. Sandri, A. Martinoni, E. Borghini, M. Civelli, G. Lamantia, S. Cinieri, G. Martinelli, C. Fiorentini, and C. M. Cipolla. 2002. Myocardial injury revealed by plasma troponin I in breast cancer treated with high-dose chemotherapy. *Annals of Oncology* 13(5):710–715.

Cardinale, D., A. Colombo, M. T. Sandri, G. Lamantia, N. Colombo, M. Civelli, G. Martinelli, F. Veglia, C. Fiorentini, and C. M. Cipolla. 2006. Prevention of high-dose chemotherapy-induced cardiotoxicity in high-risk patients by angiotensin-converting enzyme inhibition. *Circulation* 114(23):2474–2481.

Carrió, I. 2008. EANM-ESR white paper on multimodality imaging. *European Journal of Nuclear Medicine and Molecular Imaging* 35(3):673.

Casas, J. P., T. Shah, A. D. Hingorani, J. Danesh, and M. B. Pepys. 2008. C-reactive protein and coronary heart disease: A critical review. *Journal of Internal Medicine* 264(4):295–314.

Castelli, W. P. 1996. Lipids, risk factors and ischaemic heart disease. *Atherosclerosis* 124(Suppl): S1–S9.

Castelli, W. P., R. J. Garrison, P. W. Wilson, R. D. Abbott, S. Kalousdian, and W. B. Kannel. 1986. Incidence of coronary heart disease and lipoprotein cholesterol levels. The Framingham Study. *Journal of the American Medical Association* 256(20):2835–2838.

Chan, P. S., B. K. Nallamothu, R. A. Hayward. 2009. Rosuvastatin in patients with elevated c-reactive protein. *New England Journal of Medicine* 350(10):1039.

Chen, Y. N., Z. R. Luo, L. J. Zeng, M. Y. Wu, Y. Z. Wu, and Z. Y. Lin. 2000. Cardiac troponin I: A marker for post-burn cardiac injury. *Annals of Clinical Biochemistry* 37(4):447–451.

Cheung, Y. C., S. C. Chen, M. Y. Su, L. C. See, S. Hsueh, H. K. Chang, Y. C. Lin, and C. S. Tsai. 2003. Monitoring the size and response of locally advanced breast cancers to neoadjuvant chemotherapy (weekly paclitaxel and epirubicin) with serial enhanced MRI. *Breast Cancer Research and Treatment* 78(1):51–58.

Choi, H. 2005. Critical issues in response evaluation on computed tomography: Lessons from the gastrointestinal stromal tumor model. *Current Oncology Reports* 7(4):307–311.

Choi, H., C. Charnsangavej, S. C. Faria, H. A. Macapinlac, M. A. Burgess, S. R. Patel, L. L. Chen, D. A. Podoloff, and R. S. Benjamin. 2007. Correlation of computed tomography and positron emission tomography in patients with metastatic gastrointestinal stromal tumor treated at a single institution with imatinib mesylate: Proposal of new computed tomography response criteria. *Journal of Clinical Oncology* 25(13):1753–1759.

Chong, E. W., T. Y. Wong, A. J. Kreis, J. A. Simpson, and R. H. Guymer. 2007. Dietary antioxidants and primary prevention of age related macular degeneration: Systematic review and meta-analysis. *BMJ* 335(7623):755.

Christen, W. G., J. E. Manson, R. J. Glynn, J. M. Gaziano, E. Y. Chew, J. E. Buring, and C. H. Hennekens. 2007. Beta carotene supplementation and age-related maculopathy in a randomized trial of US physicians. *Archives of Ophthalmology* 125(3):333–339.

Christensen, E. 2008. Choosing the best endpoint. *Journal of Hepatology* 49(4):672–673.

Citron, M. L. 2004. Dose density in adjuvant chemotherapy for breast cancer. *Cancer Investigation* 22(4):555–568.

Cohn, E. J., L. E. Strong, W. C. Hughes, D. J. Mulford, J. N. Ashworth, M. Melin, and H. L. Taylor. 1946. Preparation and properties of serum and plasma proteolyproteins. IV. A system for the separation into fractions of the proteins lipoprotein components of biological tissues and fluids. *Journal of the American Chemical Society* 68:459–475.

Collins, R., J. Armitage, S. Parish, P. Sleight, and R. Peto. 2002. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20,536 high-risk individuals: A randomised placebo-controlled trial. *Lancet* 360(9326):7–22.

Contois, J. H., J. P. McConnell, A. A. Sethi, G. Csako, S. Devaraj, D. M. Hoefner, G. R. Warnick, and AACC Lipoproteins and Vascular Diseases Division Working Group on Best Practices. 2009. Apolipoprotein B and cardiovascular disease risk: Position statement from the AACC Lipoproteins and Vascular Diseases Division Working Group on Best Practices. *Clinical Chemistry* 55(3):407–419.

Cook, N. R. 2007. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 115(7):928–935.

Cook, N. R. 2008. Statistical evaluation of prognostic versus diagnostic models: Beyond the ROC curve. *Clinical Chemistry* 54(1):17–23.

Cook, N. R., and P. M. Ridker. 2009. Advances in measuring the effect of individual predictors of cardiovascular risk: The role of reclassification measures. *Annals of Internal Medicine* 150(11):795–802.

Cook, N. R., J. E. Buring, and P. M. Ridker. 2006. The effect of including C-reactive protein in cardiovascular risk prediction models for women. *Annals of Internal Medicine* 145(1):21–29.

Cook, N. R., C. M. Albert, J. M. Gaziano, E. Zaharris, J. MacFadyen, E. Danielson, J. E. Buring, and J. E. Manson. 2007. A randomized factorial trial of vitamins C, E, and beta-carotene in the secondary prevention of cardiovascular events in women: Results from the Women's Antioxidant Cardiovascular Study (WACS). *Archives of Internal Medicine* 167(15):1610–1618.

The Coronary Drug Project Research Group. 1975. Clofibrate and niacin in coronary heart disease. *Journal of the American Medical Association* 231(4):360–381.

Correia, L. C., M. S. Rocha, and J. P. Esteves. 2009. HDL-cholesterol level provides additional prognosis in acute coronary syndromes. *International Journal of Cardiology* 136(3):307–314.

Cromwell, W. C., J. D. Otvos, M. J. Keyes, M. J. Pencina, L. Sullivan, R. S. Vasan, P. W. Wilson, and R. B. D'Agostino. 2007. LDL particle number and risk of future cardiovascular disease in the Framingham Offspring Study—Implications for LDL management. *Journal of Clinical Lipidology* 1(6):583–592.

Danesh, J., R. Collins, P. Appleby, and R. Peto. 1998. Association of fibrinogen, C-reactive protein, albumin, or leukocyte count with coronary heart disease: Meta-analyses of prospective studies. *Journal of the American Medical Association* 279(18):1477–1482.

Danesh, J., P. Whincup, M. Walker, L. Lennon, A. Thomson, P. Appleby, J. R. Gallimore, and M. B. Pepys. 2000. Low grade inflammation and coronary heart disease: Prospective study and updated meta-analyses. *British Medical Journal* 321(7255):199–204.

Danesh, J., J. G. Wheeler, G. M. Hirschfield, S. Eda, G. Eiriksdottir, A. Rumley, G. D. O. Lowe, M. B. Pepys, and V. Gudnason. 2004. C-reactive protein and other circulating markers of inflammation in the prediction of coronary heart disease. *New England Journal of Medicine* 350(14):1387–1397.

Daniels, L. B., G. A. Laughlin, P. Clopton, A. S. Maisel, and E. Barrett-Connor. 2008a. Minimally elevated cardiac troponin T and elevated N-terminal pro-B-type natriuretic peptide predict mortality in older adults. Results from the Rancho Bernardo Study. *Journal of the American College of Cardiology* 52(6):450–459.

Daniels, L. B., G. A. Laughlin, M. J. Sarno, R. Bettencourt, R. L. Wolfert, and E. Barrett-Connor. 2008b. Lipoprotein-associated phospholipase A2 is an independent predictor of incident coronary heart disease in an apparently healthy older population. The Rancho Bernardo Study. *Journal of the American College of Cardiology* 51(9):913–919.

de Beer, F. C., A. K. Soutar, M. L. Baltz, I. Trayner, A. Feinstein, and M. B. Pepys. 1982. Low density and very low density lipoproteins are selectively bound by aggregated C-reactive protein. *Journal of Experimental Medicine* 156(1):230–242.

DeFilippi, C., S. Wasserman, S. Rosanio, E. Tiblier, H. Sperger, M. Tocchi, R. Christenson, B. Uretsky, M. Smiley, J. Gold, H. Muniz, J. Badalamenti, C. Herzog, and W. Henrich. 2003. Cardiac troponin T and C-reactive protein for predicting prognosis, coronary atherosclerosis, and cardiomyopathy in patients undergoing long-term hemodialysis. *Journal of the American Medical Association* 290(3):353–359.

Desar, I. M., C. M. van Herpen, H. W. van Laarhoven, J. O. Barentsz, W. J. Oyen, and W. T. van der Graaf. 2009. Beyond RECIST: Molecular and functional imaging techniques for evaluation of response to targeted therapy. *Cancer Treatment Reviews* 35(4):309–321.

Dhani, N., D. Tu, D. J. Sargent, L. Seymour, and M. J. Moore. 2009. Alternate endpoints for screening phase II studies. *Clinical Cancer Research* 15(6):1873–1882.

Dose Schwarz, J., M. Bader, L. Jenicke, G. Hemminger, F. Jänicke, and N. Avril. 2005. Early prediction of response to chemotherapy in metastatic breast cancer using sequential 18F-FDG PET. *Journal of Nuclear Medicine* 46(7):1144–1150.

Downs, J. R., M. Clearfield, S. Weis, E. Whitney, D. R. Shapiro, P. A. Beere, A. Langendorfer, E. A. Stein, W. Kruyer, and A. M. Gotto. 1998. Primary prevention of acute coronary events with lovastatin in men and women with average cholesterol levels. *Journal of the American Medical Association* 279:1615–1622.

Durrington, P. N. 1998. Triglycerides are more important in atherosclerosis than epidemiology has suggested. *Atherosclerosis* 141(Suppl 1):S57–S62.

Eggers, K. M., L. Lind, H. Ahlström, T. Bjerner, C. Ebeling Barbier, A. Larsson, P. Venge, and B. Lindahl. 2008. Prevalence and pathophysiological mechanisms of elevated cardiac troponin I levels in a population-based sample of elderly subjects. *European Heart Journal* 29(18):2252–2258.

Eisenhauer, E. A., P. Therasse, J. Bogaerts, L. H. Schwartz, D. Sargent, R. Ford, J. Dancey, S. Arbuck, S. Gwyther, M. Mooney, L. Rubinstein, L. Shankar, L. Dodd, R. Kaplan, D. Lacombe, and J. Verweij. 2009. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *European Journal of Cancer* 45(2):228–247.

El Harchaoui, K., W. A. van der Steeg, E. S. G. Stroes, J. A. Kuivenhoven, J. D. Otvos, N. J. Wareham, B. A. Hutten, J. J. P. Kastelein, K.-T. Khaw, and S. M. Boekholdt. 2007. Value of low-density lipoprotein particle number and size as predictors of coronary artery disease in apparently healthy men and women: The EPIC-Norfolk Prospective Population Study. *Journal of the American College of Cardiology* 49(5):547–553.

Elliott, P., J. C. Chambers, W. Zhang, R. Clarke, J. C. Hopewell, J. F. Peden, J. Erdmann, P. Braund, J. C. Engert, D. Bennett, L. Coin, D. Ashby, I. Tzoulaki, I. J. Brown, S. Mt-Isa, M. I. McCarthy, L. Peltonen, N. B. Freimer, M. Farrall, A. Ruokonen, A. Hamsten, N. Lim, P. Froguel, D. M. Waterworth, P. Vollenweider, G. Waeber, M.-R. Jarvelin, V. Mooser, J. Scott, A. S. Hall, H. Schunkert, S. S. Anand, R. Collins, N. J. Samani, H. Watkins, and J. S. Kooner. 2009. Genetic loci associated with C-reactive protein levels and risk of coronary heart disease. *Journal of the American Medical Association* 302(1):37–48.

Eng-Wong, J., J. Orzano-Birgani, C. K. Chow, D. Venzon, J. Yao, C. E. Galbo, J. A. Zujewski, and S. Prindiville. 2008. Effect of raloxifene on mammographic density and breast magnetic resonance imaging in premenopausal women at increased risk for breast cancer. *Cancer Epidemiology, Biomarkers, and Prevention* 17(7):1696–1701.

Feeney, L., and E. R. Berman. 1976. Oxygen toxicity: Membrane damage by free radicals. *Investigative Ophthalmology* 15(10):789–792.

Feenstra, J., B. A. in't Veld, P. D. van der Linden, D. E. Grobbee, and B. H. Stricker. 1998. Risk factors for mortality in users of ibopamine. *British Journal of Clinical Pharmacology* 46(1):71–77.

FHS (Framingham Heart Study). 2009. *Framingham Heart Study: A project of the National Heart, Lung and Blood Institute and Boston University*. http://www.framinghamheartstudy.org/about/index.html (accessed October 20, 2009).

Fishman, E. K., E. Mendelson, M. Böhm-Vélez, R. Bree, H. Finberg, H. Hricak, F. Laing, D. Sartoris, A. Thurmond, S. Goldstein, and J. Walsh. 2000. Staging and follow-up of ovarian cancer. American College of Radiology. ACR Appropriateness Criteria. *Radiology* 215(Suppl):899–902.

Fleming, S. M., T. O'Gorman, J. Finn, H. Grimes, K. Daly, and J. J. Morrison. 2000. Cardiac troponin I in pre-eclampsia and gestational hypertension. *British Journal of Obstetrics and Gynaecology* 107(11):1417–1420.

Fleming, T. R., M. D. Rothmann, and H. L. Lu. 2009. Issues in using progression-free survival when evaluating oncology products. *Journal of Clinical Oncology* 27(17):2874–2880.

Folsom, A. R., L. E. Chambless, C. M. Ballantyne, J. Coresh, G. Heiss, K. K. Wu, E. Boerwinkle, T. H. Mosley, P. Sorlie, G. Diao, and A. R. Sharrett. 2006. An assessment of incremental coronary risk prediction using C-reactive protein and other novel risk markers: The Atherosclerosis Risk in Communities Study. *Archives of Internal Medicine* 166(13):1368–1373.

Foot, D. K., R. P. Lewis, T. A. Pearson, and G. A. Beller. 2000. Demographics and cardiology, 1950–2050. *Journal of the American College of Cardiology* 35(5 Suppl B):66B–80B.

Frick, M. H., O. Elo, K. Haapa, O. P. Heinonen, P. Heinsalmi, P. Helo, J. K. Huttunen, P. Kaitaniemi, P. Koskinen, V. Manninen et al. 1987. Helsinki Heart Study: Primary-prevention trial with gemfibrozil in middle-aged men with dyslipidemia. Safety of treatment, changes in risk factors, and incidence of coronary heart disease. *New England Journal of Medicine* 317(20):1237–1245.

Funaioli, C., C. Pinto, F. Di Fabio, D. Santini, C. Ceccarelli, E. De Raffaele, S. Fanti, P. Castellucci, C. Longobardi, F. Buggi, and A. A. Martoni. 2007. 18FDG-PET evaluation correlates better than CT with pathological response in a metastatic colon cancer patient treated with bevacizumab-based therapy. *Tumori* 93(6):611–615.

Gaines Das, R. E., and S. Poole. 1993. The international standard for interleukin-6. Evaluation in an international collaborative study. *Journal of Immunological Methods* 160(2):147–153.

Galanis, E., J. C. Buckner, M. J. Maurer, R. Sykora, R. Castillo, K. V. Ballman, and B. J. Erickson. 2006. Validation of neuroradiologic response assessment in gliomas: Measurement by RECIST, two-dimensional, computer-assisted tumor area, and computer-assisted tumor volume methods. *Neuro-Oncology* 8(2):156–165.

Gehan, E. A., and M. C. Tefft. 2000. Will there be resistance to the RECIST (Response Evaluation Criteria in Solid Tumors)? *Journal of the National Cancer Institute* 92(3):179–181.

Gofman, J. W., and F. Lindgren. 1950. The role of lipids and lipoproteins in atherosclerosis. *Science* 111(2877):166–171.

Gofman, J. W., H. B. Jones, F. T. Lindgren, T. P. Lyon, H. A. Elliott, and B. Strisower. 1950a. Blood lipids and human atherosclerosis. *Circulation* 2(2):161–178.

Gofman, J. W., F. T. Lindgren, H. B. Jones, T. P. Lyon, and B. Strisower. 1950b. Lipoproteins and atherosclerosis. *Journal of Gerontology* 6:105–119.

Gofman, J., O. De Lalla, and F. F. Glazier, N. K. Lindgren, F. T. 1954. The serum lipoprotein transport system in health, metabolic disorders, atherosclerosis, and coronary heart disease. *Plasma* 2:413–484.

Goldberg, J., G. Flowerdew, E. Smith, J. A. Brody, and M. O. Tso. 1988. Factors associated with age-related macular degeneration: Analysis of data from the first National Health and Nutrition Examination Survey. *American Journal of Epidemiology* 128(4):700–710.

Goldstein, J. L., W. R. Hazzard, H. G. Schrott, E. L. Bierman, and A. G. Motulsky. 1973. Hyperlipidemia in coronary heart disease: Lipid levels in 500 survivors of myocardial infarction. *Journal of Clinical Investigation* 52:1533–1543.

Goldstein, D., B. S. Tan, M. Rossleigh, W. Haindl, B. Walker, and J. Dixon. 2005. Gastrointestinal stromal tumours: Correlation of F-FDG gamma camera-based coincidence positron emission tomography with CT for the assessment of treatment response—An AGITG study. *Oncology* 69(4):326–332.

Gordon, D. J., and B. M. Rifkind. 1989. High-density lipoprotein—The clinical implications of recent studies. *New England Journal of Medicine* 321(19):1311–1316.

Gorecki, A., B. Bednarz, T. Jaxa-Chamiec, P. Maciejewski, R. Lukaszewicz, L. Ceremuzynski, and A. Dyduszynski. 2004. Lipid profile during the first 24 hours after myocardial infarction has significant prognostic value. *Kardiologia Polska* 60(3):229–236, 237.

Greenland, P., M. D. Knoll, J. Stamler, J. D. Neaton, A. R. Dyer, D. B. Garside, P. W. Wilson. 2003. Major risk factors as antecedents of fatal and nonfatal coronary heart disease events. *Journal of the American Medical Association* 290(7):891-897.

Griffin, J. H., K. Kojima, C. L. Banka, L. K. Curtiss, and J. A. Fernández. 1999. High-density lipoprotein enhancement of anticoagulant activities of plasma protein S and activated protein C. *Journal of Clinical Investigation* 103(2):219–227.

Gritz, D. C., M. Srinivasan, S. D. Smith, U. Kim, T. M. Lietman, J. H. Wilkins, B. Priyadharshini, R. K. John, S. Aravind, N. V. Prajna, R. Duraisami Thulasiraj, and J. P. Whitcher. 2006. The Antioxidants in Prevention of Cataracts Study: Effects of antioxidant supplements on cataract progression in South India. *British Journal of Ophthalmology* 90(7):847–851.

Grossi, F., O. Belvedere, G. Fasola, T. Ceschia, S. Meduri, A. Sibau, C. Sacco, M. Bazzocchi, G. Talmassons, A. Morelli, F. Barbone, and A. Sobrero. 2004. Tumor measurements on computed tomographic images of non-small cell lung cancer were similar among cancer professionals from different specialties. *Journal of Clinical Epidemiology* 57(8):804–808.

Guest, T. M., A. V. Ramanathan, P. G. Tuteur, K. B. Schechtman, J. H. Ladenson, and A. S. Jaffe. 1995. Myocardial injury in critically ill patients: A frequently unrecognized complication. *Journal of the American Medical Association* 273(24):1945–1949.

Hamm, C. W., and E. Braunwald. 2000. A classification of unstable angina revisited. *Circulation* 102(1):118–122.

Havel, R. J., H. A. Eder, and J. H. Bragdon. 1955. The distribution and chemical composition of ultracentrifugally separated lipoproteins in human serum. *The Journal of Clinical Investigation* 34:1345–1353.

Hayward, R. A., T. P. Hofer, and S. Vijan. 2006. Narrative review: Lack of evidence for recommended low-density lipoprotein treatment targets: A solvable problem. *Annals of Internal Medicine* 145(7):520-530.

Heidenreich, P. A., T. Alloggiamento, K. Melsop, K. M. McDonald, A. S. Go, and M. A. Hlatky. 2001. The prognostic value of troponin in patients with non-ST elevation acute coronary syndromes: A meta-analysis. *Journal of the American College of Cardiology* 38(2):478–485.

Heinonen, O. P., and D. Albanes. 1994. The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. *New England Journal of Medicine* 330(15):1029–1035.

Heliovaara, M., P. Knekt, K. Aho, R. K. Aaran, G. Alfthan, and A. Aromaa. 1994. Serum antioxidants and risk of rheumatoid arthritis. *Annals of the Rheumatic Diseases* 53(1):51–53.

Hennekens, C., J. Buring, J. Manson, M. Stampfer, B. Rosner, N. Cook, C. Belanger, F. LaMotte, J. M. Gaziano, P. Ridker, W. Willett, and R. Peto. 1996. Lack of effect of long-term supplementation with beta-carotene on the incidence of malignant neoplasms and cardiovascular disease. *New England Journal of Medicine* 334(18):1145–1149.

Henson, J. W., P. Gaviani, and R. G. Gonzalez. 2005. MRI in treatment of adult gliomas. *Lancet Oncology* 6(3):167–175.

Herrington, D. M., D. M. Reboussin, K. B. Brosnihan, P. C. Sharp, S. A. Shumaker, T. E. Snyder, C. D. Furberg, G. J. Kowalchuk, T. D. Stuckey, W. J. Rogers, D. H. Givens, and D. Waters. 2000. Effects of estrogen replacement on the progression of coronary-artery atherosclerosis. *New England Journal of Medicine* 343(8):522–529.

HHS (Department of Health and Human Services). 1990. *The health benefits of smoking cessation: A report of the Surgeon General.* Rockville, MD: HHS.

HHS. 1996. Clinton administration announces plan to speed development and approval of cancer drugs. http://www.hhs.gov/news/press/1996pres/960329.html (accessed February 17, 2010).

Hirschfield, G. M., R. Gallimore, J. A. Gilbertson, M. Benson, and M. B. Pepys. 2003. Systemic inflammation and atherogenesis in apoE knockout mice expressing transgenic human c-reactive protein. *Circulation* 108(Suppl IV):IV-41. Abstract.

Hirschfield, G. M., J. R. Gallimore, M. C. Kahan, W. L. Hutchinson, C. A. Sabin, G. M. Benson, A. P. Dhillon, G. A. Tennent, and M. B. Pepys. 2005. Transgenic human c-reactive protein is not proatherogenic in apolipoprotein E-deficient mice. *Proceedings of the National Academy of Sciences of the United States of America* 102(23): 8309-8314.

Hlatky, M. A. 2008. Expanding the orbit of primary prevention—Moving beyond JUPITER. *New England Journal of Medicine* 359(21):2280–2282.

Hoefner, D. M. 2008. The cholesterol controversy continues. *MLO: Medical Laboratory Observer* 40(4):24, 26, 28 passim.

Hohenberger, P., and E. Wardelmann. 2006. [Surgical considerations for gastrointestinal stroma tumor]. *Der Chirurg* 77(1):33–40.

Hopkins, P. N., and R. R. Williams. 1981. A survey of 246 suggested coronary risk factors. *Atherosclerosis* 40(1):1–52.

Horne, B. D., J. B. Muhlestein, J. F. Carlquist, T. L. Bair, T. E. Madsen, N. I. Hart, and J. L. Anderson. 2000. Statin therapy, lipid levels, C-reactive protein and the survival of patients with angiographically severe coronary artery disease. *Journal of the American College of Cardiology* 36(6):1774–1780.

Hudis, C. A. 2005. Clinical implications of antiangiogenic therapies. *Oncology* 19(4 Suppl 3):26–31.

Hulley, S., D. Grady, T. Bush, C. Furberg, D. Herrington, B. Riggs, E. Vittinghoff, and for the Heart and Estrogen/progestin Replacement Study Research Group. 1998. Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. *Journal of the American Medical Association* 280(7):605–613.

Husband, J. E., L. H. Schwartz, J. Spencer, L. Ollivier, D. M. King, R. Johnson, and R. Reznek. 2004. Evaluation of the response to treatment of solid tumours—A consensus statement of the International Cancer Imaging Society. *British Journal of Cancer* 90(12):2256–2260.

Hylton, N. 2006. MR imaging for assessment of breast cancer response to neoadjuvant chemotherapy. *Magnetic Resonance Imaging Clinics of North America* 14(3):383–389, vii.

Imamoto, A., L. M. Beltrán, and J. DiGiovanni. 1991. Evidence for autocrine/paracrine growth stimulation by transforming growth factor-alpha during the process of skin tumor promotion. *Molecular Carcinogenesis* 4(1):52–60.

Inazu, A., M. L. Brown, C. B. Hesler, L. B. Agellon, J. Koizumi, K. Takata, Y. Muruhama, H. Mabuchi, and A. R. Tall. 1990. Increased high-density lipoprotein levels caused by a common cholesteryl-ester transfer protein gene mutation. *New England Journal of Medicine* 323(18):1234–1238.

IOM (Institute of Medicine). 2000a. *Dietary Reference Intakes for vitamin A, vitamin K, arsenic, boron, chromium, copper, iodine, iron, manganese, molybdenum, nickel, silicon, vanadium, and zinc*. Washington, DC: National Academy Press.

IOM. 2000b. *Dietary Reference Intakes for vitamin C, vitamin E, selenium, and carotenoids*. Washington, DC: National Academy Press.

IOM. 2007. *Cancer biomarkers: The promises and challenges of improving detection and treatment*. Washington, DC: The National Academies Press.

IOM. 2009. *Accelerating the development of biomarkers for drug safety*. Washington, DC: The National Academies Press.

Jaffe, A. S. 2001. Elevations in cardiac troponin measurements: False false-positives: The real truth. *Cardiovascular Toxicology* 1(2):87–92.

Jaffe, A. S., J. Ravkilde, R. Roberts, U. Naslund, F. S. Apple, M. Galvani, and H. Katus. 2000. It's time for a change to a troponin standard. *Circulation* 102(11):1216–1220.

James, P., C. J. Ellis, R. M. L. Whitlock, A. R. McNeil, J. Henley, and N. E. Anderson. 2000. Relation between troponin T concentration and mortality in patients presenting with an acute stroke: Observational study. *British Medical Journal* 320(7248):1502–1504.

Javitt, M. C. 2007. ACR Appropriateness Criteria on staging and follow-up of ovarian cancer. *Journal of the American College of Radiology* 4(9):586–589.

Jennings, D., N. Raghunand, and R. J. Gillies. 2008. Imaging hemodynamics. *Cancer Metastasis Reviews* 27(4):589–613.

Jensen, M. M., J. T. Jørgensen, T. Binderup, and A. Kjaer. 2008. Tumor volume in subcutaneous mouse xenografts measured by microCT is more accurate and reproducible than determined by 18F-FDG-microPET or external caliper. *BMC Medical Imaging* 8:16.

Jiang, X. C., L. B. Agellon, A. Walsh, J. L. Breslow, and A. Tall. 1992. Dietary cholesterol increases transcription of the human cholesteryl ester transfer protein gene in transgenic mice. Dependence on natural flanking sequences. *Journal of Clinical Investigation* 90(4):1290–1295.

Jiang, R., M. B. Schulze, T. Li, N. Rifai, M. J. Stampfer, E. B. Rimm, and F. B. Hu. 2004. Non-HDL cholesterol and apolipoprotein B predict cardiovascular disease events among men with type 2 diabetes. *Diabetes Care* 27(8):1991–1997.

Joy, T., and R. A. Hegele. 2008a. Is raising HDL a futile strategy for atheroprotection? *Nature Reviews Drug Discovery* 7(2):143–155.

Joy, T. R., and R. A. Hegele. 2008b. The failure of torcetrapib: What have we learned? *British Journal of Pharmacology* 154(7):1379–1381.

Kamel, I. R., and D. A. Bluemke. 2002. Magnetic resonance imaging of the liver: Assessing response to treatment. *Topics Magnetic Resonance Imaging* 13(3):191–200.

Kannel, W. B., T. R. Dawber, A. Kagan, N. Revotskie, and J. I. Stokes. 1961. Factors of risk in the development of coronary heart disease—six-year follow-up experience. The Framingham Heart Study. *Annals of Internal Medicine* 55(1):33–50.

Kaplan, I. V., and S. S. Levinson. 1999. When is a heterophile antibody not a heterophile antibody? When it is an antibody against a specific immunogen. *Clinical Chemistry* 45(5):616–618.

Karrison, T. G., M. L. Maitland, W. M. Stadler, and M. J. Ratain. 2007. Design of phase II cancer trials using a continuous endpoint of change in tumor size: Application to a study of sorafenib and erlotinib in non small-cell lung cancer. *Journal of the National Cancer Institute* 99(19):1455–1461.

Katritsis, D. G., M. Hossein-Nia, A. Anastasakis, J. Poloniecki, D. W. Holt, A. J. Camm, D. E. Ward, and E. Rowland. 1998. Myocardial injury induced by radiofrequency and low energy ablation: A quantitative study of CK isoforms, CK-MB, and troponin-T concentrations. *PACE - Pacing and Clinical Electrophysiology* 21(7):1410–1416.

Katrukha, A. G., A. V. Bereznikova, V. L. Filatov, T. V. Esakova, O. V. Kolosova, K. Pettersson, T. Lövgren, T. V. Bulargina, I. R. Trifonov, N. A. Gratsiansky, K. Pulkki, L. M. Voipio-Pulkki, and N. B. Gusev. 1998. Degradation of cardiac troponin I: Implication for reliable immunodetection. *Clinical Chemistry* 44(12):2433–2440.

Kerbel, R. S. 2006. Antiangiogenic therapy: A universal chemosensitization strategy for cancer? *Science* 312(5777):1171–1175.

Khot, U. N., M. B. Khot, C. T. Bajzer, S. K. Sapp, E. M. Ohman, S. J. Brener, S. G. Ellis, A. M. Lincoff, and E. J. Topol. 2003. Prevalence of conventional risk factors in patients with coronary heart disease. *Journal of the American Medical Association* 290(7):898–904.

Kimberly, M. M., H. W. Vesper, S. P. Caudill, G. R. Cooper, N. Rifai, F. Dati, and G. L. Myers. 2003. Standardization of immunoassays for measurement of high-sensitivity C-reactive protein. Phase I: Evaluation of secondary reference materials. *Clinical Chemistry* 49(4):611–616.

Kimberly, M. M., S. P. Caudill, H. W. Vesper, E. A. Monsell, W. G. Miller, R. Rej, N. Rifai, F. Dati, and G. L. Myers. 2009. Standardization of high-sensitivity immunoassays for measurement of C-reactive protein; II: Two approaches for assessing commutability of a reference material. *Clinical Chemistry* 55(2):342–350.

Kneckt, P., M. Heliovaara, A. Rissanen, A. Aromaa, and R. K. Aaran. 1992. Serum antioxidant vitamins and risk of cataract. *BMJ* 305(6866):1392–1394.

Knopp, R. H., P. Paramsothy, B. M. Retzlaff, B. Fish, C. Walden, A. Dowdy, C. Tsunehara, K. Aikawa, and M. C. Cheung. 2005. Gender differences in lipoprotein metabolism and dietary response: Basis in hormonal differences and implications for cardiovascular disease. *Current Atherosclerosis Reports* 7(6):472–479.

Knopp, R. H., P. Paramsothy, B. Atkinson, and A. Dowdy. 2008. Comprehensive lipid management versus aggressive low-density lipoprotein lowering to reduce cardiovascular risk. *American Journal of Cardiology* 101(8A):48B–57B.

Kolata, G. 2009. Study dismisses protein's role in heart disease. *The New York Times*, July 1, A18.

Krauss, R. M. 1995. Dense low density lipoproteins and coronary artery disease. *American Journal of Cardiology* 75(6):53B–57B.

Krauss, R. M., and D. J. Burke. 1982. Identification of multiple subclasses of plasma low density lipoproteins in normal humans. *Journal of Lipid Research* 23(1):97–104.

Kronmal, R. A., K. C. Cain, Z. Ye, and G. S. Omenn. 1993. Total serum cholesterol levels and mortality risk as a function of age: A report based on the Framingham data. *Archives of Internal Medicine* 153(9):1065–1073.

Krumholz, H. M., and T. H. Lee. 2008. Redefining quality—Implications of recent clinical trials. *New England Journal of Medicine* 358(24):2537–2539.

LaRosa, J. C., S. M. Grundy, D. D. Waters, C. Shear, P. Barter, J.-C. Fruchart, A. M. Gotto, H. Greten, J. J. P. Kastelein, J. Shepherd, N. K. Wenger, and the Treating to New Targets Investigators. 2005. Intensive lipid lowering with atorvastatin in patients with stable coronary disease. *New England Journal of Medicine* 352(14):1425–1435.

Lathia, C. D., D. Amakye, W. Dai, C. Girman, S. Madani, J. Mayne, P. MacCarthy, P. Pertel, L. Seman, A. Stoch, P. Tarantino, C. Webster, S. Williams, and J. A. Wagner. 2009. The value, qualification, and regulatory use of surrogate end points in drug development. *Clinical Pharmacology and Therapeutics* 86(1):32–43.

Latini, R., S. Masson, I. S. Anand, E. Missov, M. Carlson, T. Vago, L. Angelici, S. Barlera, G. Parrinello, A. P. Maggioni, G. Tognoni, and J. N. Cohn. 2007. Prognostic value of very low plasma concentrations of troponin T in patients with stable chronic heart failure. *Circulation* 116(11):1242–1249.

Lauer, B., C. Niederau, U. Kühl, M. Schannwell, M. Pauschinger, B. E. Strauer, and H. P. Schultheiss. 1997. Cardiac troponin T in patients with clinically suspected myocarditis. *Journal of the American College of Cardiology* 30(5):1354–1359.

Ledue, T. B., and N. Rifai. 2003. Preanalytic and analytic sources of variations in C-reactive protein measurement: Implications for cardiovascular disease risk assessment. *Clinical Chemistry* 49(8):1258–1271.

Lee, I. M., N. R. Cook, J. E. Manson, J. E. Buring, and C. H. Hennekens. 1999. Beta-carotene supplementation and incidence of cancer and cardiovascular disease: The Women's Health Study. *Journal of the National Cancer Institute* 91(24):2102–2106.

Leimgruber, A., S. Ostermann, E. J. Yeon, E. Buff, P. P. Maeder, R. Stupp, and R. A. Meuli. 2006. Perfusion and diffusion MRI of glioblastoma progression in a four-year prospective temozolomide clinical trial. *International Journal of Radiation Oncology, Biology, Physics* 64(3):869–875.

Liao, J. K., and U. Laufs. 2005. Pleiotropic effects of statins. *Annual Review of Pharmacology and Toxicology* 45:89–118.

Libby, P., and P. M. Ridker. 1999. Novel inflammatory markers of coronary risk: Theory versus practice. *Circulation* 100(11):1148–1150.

Libby, P., and P. Theroux. 2005. Pathophysiology of coronary artery disease. *Circulation* 111(25):3481–3488.

Lichtenstein, A. V. 2005. Cancer as a programmed death of an organism. *Biochemistry (Mosc)* 70(9):1055–1064.

Lin, N. U., L. A. Carey, M. C. Liu, J. Younger, S. E. Come, M. Ewend, G. J. Harris, E. Bullitt, A. D. Van den Abbeele, J. W. Henson, X. Li, R. Gelman, H. J. Burstein, E. Kasparian, D. G. Kirsch, A. Crawford, F. Hochberg, and E. P. Winer. 2008. Phase II trial of lapatinib for brain metastases in patients with human epidermal growth factor receptor 2-positive breast cancer. *Journal of Clinical Oncology* 26(12):1993–1999.

Lipid Research Clinics Program. 1984a. The lipid research clinics coronary primary prevention trial results I. Reduction in incidence of coronary heart disease. *Journal of the American Medical Association* 251(3):351–364.

Lipid Research Clinics Program. 1984b. The lipid research clinics coronary primary prevention trial results II. The relationship of reduction in incidence of coronary heart disease to cholesterol lowering. *Journal of the American Medical Association* 251(3):365–374.

Lipid Research Clinics Program. 1984c. The relationship of reduction in incidence of coronary heart disease to cholesterol lowering. *Journal of the American Medical Association* 252(18):2545–2548.

Liu, S., I. M. Lee, U. Ajani, S. R. Cole, J. E. Buring, J. E. Manson, and Physicians' Health Study. 2001. Intake of vegetables rich in carotenoids and risk of coronary heart disease in men: The Physicians' Health Study. *International Journal of Epidemiology* 30(1):130–135.

Llovet, J. M., A. M. Di Bisceglie, J. Bruix, B. S. Kramer, R. Lencioni, A. X. Zhu, M. Sherman, M. Schwartz, M. Lotze, J. Talwalkar, and G. J. Gores. 2008. Design and endpoints of clinical trials in hepatocellular carcinoma. *Journal of the National Cancer Institute* 100(10):698–711.

Lloyd-Jones, D., R. Adams, M. Carnethon, G. De Simone, T. B. Ferguson, K. Flegal, E. Ford, K. Furie, A. Go, K. Greenlund, N. Haase, S. Hailpern, M. Ho, V. Howard, B. Kissela, S. Kittner, D. Lackland, L. Lisabeth, A. Marelli, M. McDermott, J. Meigs, D. Mozaffarian, G. P. Sorlie, R. Stafford, J. Steinberger, T. Thom, S. Wasserthiel-Smoller, N. Wong, J. Wylie-Rosett, and Y. Hong. 2009. Heart disease and stroke statistics 2009 update: A report from the American Heart Association Statistics Committee and Stroke Statistics Subcomittee. *Circulation* 119(3):e21–e181.

Lopes-Virella, M. F., and G. Virella. 2003. The role of immune and inflammatory processes in the development of macrovascular disease in diabetes. *Frontiers in Bioscience* 8: s750–s768.

Macheboeuf, M. 1929. Recherches sur les phosphoaminolipides et les sterids du serum et du plasma sanguins. II Etude physiochimique de la fraction proteidique la plus riche en phospholipids et in sterides. *Bulletin de la Société de chimie biologique* 11:485–503.

MacMahon, S., R. Peto, R. Collins, J. Godwin, S. MacMahon, J. Cutler, P. Sorlie, R. Abbott, R. Collins, J. Neaton, R. Abbott, A. Dyer, and J. Stamler. 1990. Blood pressure, stroke, and coronary heart disease, I: Prolonged differences in blood pressure: Prospective observational studies corrected for the regression dilution bias. *Lancet* 335(8692):765–774.

Manninen, V., L. Tenkanen, P. Koskinen, J. K. Huttunen, M. Manttari, O. P. Heinonen, and M. H. Frick. 1992. Joint effects of serum triglyceride and LDL cholesterol and HDL cholesterol concentrations on coronary heart disease risk in the Helsinki Heart Study. Implications for treatment. *Circulation* 85(1):37–45.

Marcus, C. D., V. Ladam-Marcus, C. Cucu, O. Bouché, L. Lucas, and C. Hoeffel. 2009. Imaging techniques to evaluate the response to treatment in oncology: Current standards and perspectives. *Critical Reviews in Oncology/Hematology* 72(3):217–238.

Marques-Vidal, P., J. Amar, J. P. Cambou, and B. Chamontin. 1996. Relationships between blood pressure components, lipids and lipoproteins in normotensive men. *Journal of Human Hypertension* 10(4):239–244.

McNamara, J. R., G. R. Warnick, and G. R. Cooper. 2006. A brief history of lipid and lipoprotein measurements and their contribution to clinical chemistry. *Clinica Chimica Acta* 369(2):158–167.

Mensah, G. A., and D. W. Brown. 2007. An overview of cardiovascular disease burden in the United States. *Health Affairs* 26(1):38–48.

Michos, E. D., and R. S. Blumenthal. 2009. Prevalence of low low-density lipoprotein cholesterol with elevated high sensitivity C-reactive protein. *Journal of the American College of Cardiology* 53(11):931–935.

Miller, W. G., P. P. Waymack, F. P. Anderson, S. F. Ethridge, and E. C. Jayne. 2002. Performance of four homogeneous direct methods for LDL-cholesterol. *Clinical Chemistry* 48(3):489–498.

Miremadi, S., A. Sniderman, and J. Frohlich. 2002. Can measurement of serum apolipoprotein B replace the lipid profile monitoring of patients with lipoprotein disorders? *Clinical Chemistry* 48(3):484–488.

Missov, E., C. Calzolari, and B. Pau. 1997. Circulating cardiac troponin I in severe congestive heart failure. *Circulation* 96(9):2953–2958.

Monteil, J., N. Mahmoudi, S. Leobon, P. Y. Roudaut, A. El Badaoui, S. Verbeke, L. Venat-Bouvet, J. Martin, V. Le Brun-Ly, S. Lavau-Denes, A. Maubon, P. Bouillet, M. Pouquet, J. C. Vandroux, and N. Tubiana-Mathieu. 2009. Chemotherapy response evaluation in metastatic colorectal cancer with FDG PET/CT and CT scans. *Anticancer Research* 29(7):2563–2568.

Mora, S., J. D. Otvos, N. Rifai, R. S. Rosenson, J. E. Buring, and P. M. Ridker. 2009a. Lipoprotein particle profiles by nuclear magnetic resonance compared with standard lipids and apolipoproteins in predicting incident cardiovascular disease in women. *Circulation* 119(7):931–939.

Mora, S., N. Rifai, J. E. Buring, and P. M. Ridker. 2009b. Comparison of LDL cholesterol concentrations by Friedewald calculation and direct measurement in relation to cardiovascular events in 27,331 women. *Clinical Chemistry* 55(5):888–894.

Morrow, D. A., C. P. Cannon, N. Rifai, M. J. Frey, R. Vicari, N. Lakkis, D. H. Robertson, D. A. Hille, P. T. DeLucca, P. M. DiBattiste, L. A. Demopoulos, W. S. Weintraub, and E. Braunwald. 2001. Ability of minor elevations of troponins I and T to predict benefit from an early invasive strategy in patients with unstable angina and non-ST elevation myocardial infarction: Results from a randomized trial. *Journal of the American Medical Association* 286(19):2405–2412.

Morrow, D. A., C. P. Cannon, R. L. Jesse, L. K. Newby, J. Ravkilde, A. B. Storrow, A. H. B. Wu, R. H. Christenson, F. S. Apple, G. Francis, and W. Tang. 2007. National Academy of Clinical Biochemistry Laboratory Medicine Practice Guidelines: Clinical characteristics and utilization of biochemical markers in acute coronary syndromes. *Clinical Chemistry* 53(4):552–574.

Musunuru, K., B. G. Kral, R. S. Blumenthal, V. Fuster, C. Y. Campbell, T. J. Gluckman, R. A. Lange, E. J. Topol, J. T. Willerson, M. Y. Desai, M. H. Davidson, and S. Mora. 2008. The use of high-sensitivity assays for C-reactive protein in clinical practice. *Nature Clinical Practice. Cardiovascular Medicine* 5(10):621–635.

Musunuru, K., M. Orho-Melander, M. P. Caulfield, S. Li, W. A. Salameh, R. E. Reitz, G. Berglund, B. Hedblad, G. Engstrom, P. T. Williams, S. Kathiresan, O. Melander, and R. M. Krauss. 2009. Ion mobility analysis of lipoprotein subfractions identifies three independent axes of cardiovascular risk. *Arteriosclerosis Thrombosis and Vascular Biology* 29(11):1975–1980.

Mykkanen, L., J. Kuusisto, S. Haffner, M. Laakso, and M. Austin. 1999. LDL size and risk of coronary heart disease in elderly men and women. *Arteriosclerosis, Thrombosis, and Vascular Biology* 19:2742–2748.

NACB Writing Group Members, A. H. B. Wu, A. S. Jaffe, F. S. Apple, R. L. Jesse, G. L. Francis, D. A. Morrow, L. K. Newby, J. Ravkilde, W. H. Wilson Tang, R. H. Christenson; NACB Committee Members, R. H. Christenson, F. S. Apple, C. P. Cannon, G. L. Francis, R. L. Jesse, D. A. Morrow, L. K. Newby, J. Ravkilde, A. B. Storrow, W. H. Wilson Tang, and A. H. B. Wu. 2007. National Academy of Clinical Biochemistry Laboratory Medicine practice guidelines: Use of cardiac troponin and B-type natriuretic peptide or N-terminal proB-type natriuretic peptide for etiologies other than acute coronary syndromes and heart failure. *Clinical Chemistry* 53(12):2086–2096.

Nauck, M., G. R. Warnick, and N. M. Rifai. 2002. Methods for measurement of LDL-cholesterol: A critical assessment of direct measurement by homogeneous assays versus calculation. *Clinical Chemistry* 48:236–254.

NCEP (National Cholesterol Education Program). 2001. *Third Report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III)*. Bethesda, MD: National Institutes of Health.

Needham, D. M., K. A. Shufelt, G. Tomlinson, J. W. Scholey, and G. E. Newton. 2004. Troponin I and T levels in renal failure patients without acute coronary syndrome: A systematic review of the literature. *Canadian Journal of Cardiology* 20(12):1212–1218.

Nicoletto, M. O., D. Nitti, L. Pescarini, F. Corbetti, R. Mencarelli, A. Cappetta, A. Galligioni, C. Pogliani, A. Marchet, F. Bozza, C. Ghiotto, L. Griggio, G. Zavagno, M. E. Donach, and C. Di Maggio. 2008. Correlation between magnetic resonance imaging and histopathological tumor response after neoadjuvant chemotherapy in breast cancer. *Tumori* 94(4):481–488.

Nilsson, J. 2005. CRP—marker or maker of cardiovascular disease? *Arterioscler Thromb Vasc Biol* 25(8):1527–1528.

Nissen, S. 2008. ENHANCE and ACCORD: Controversy over surrogate end points. *Current Cardiology Reports* 10(3):159–161.

Nissen, S. E., T. Tsunoda, E. M. Tuzcu, P. Schoenhagen, C. J. Cooper, M. Yasin, G. M. Eaton, M. A. Lauer, W. S. Sheldon, C. L. Grines, S. Halpern, T. Crowe, J. C. Blankenship, and R. Kerensky. 2003. Effect of recombinant ApoA-I Milano on coronary atherosclerosis in patients with acute coronary syndromes: A randomized controlled trial. *Journal of the American Medical Association* 290(17):2292–2300.

Nofer, J. R., B. Kehrel, M. Fobker, B. Levkau, G. Assmann, and A. von Eckardstein. 2002. HDL and arteriosclerosis: Beyond reverse cholesterol transport. *Atherosclerosis* 161(1):1–16.

Nordestgaard, B. G. 2009. Does elevated C-reactive protein cause human atherothrombosis? Novel insights from genetics, intervention trials, and elsewhere. *Current Opinion in Lipidology* 20(5):393–401.

Noterdaeme, O., M. Kelly, P. Friend, Z. Soonowalla, G. Steers, and M. Brady. 2009. Initial assessment of a model relating intratumoral genetic heterogeneity to radiological morphology. *British Journal of Radiology* 83(986):166–170.

Ogbuanu, I. U., H. Zhang, and W. Karmaus. 2009. Can we apply the Mendelian randomization methodology without considering epigenetic effects? *Emerging Themes in Epidemiology* 6(3):e1–e10.

Ohara, K., Y. Tanaka, H. Tsunoda, M. Nishida, S. Sugahara, and Y. Itai. 2002. Assessment of cervical cancer radioresponse by serum squamous cell carcinoma antigen and magnetic resonance imaging. *Obstetrics and Gynecology* 100(4):781–787.

Ollivier, L., J. Leclère, P. Thiesse, D. Di Stefano, and C. Vincent. 2007. [Measurement of tumour response to cancer treatment: Morphologic imaging role]. *Bulletin du Cancer* 94(2):171–177.

Olson, R. E. 1998. Discovery of the lipoproteins, their role in fat transport and their significance as risk factors. *Journal of Nutrition* 128(2 Suppl):439S–443S.

Omenn, G. S., G. E. Goodman, M. D. Thornquist, J. Balmes, M. R. Cullen, A. Glass, J. P. Keogh, J. F. L. Meyskens Jr., B. Valanis, J. H. Williams Jr., S. Barnhart, and S. Hammar. 1996. Effects of a combination of beta-carotene and vitamin A on lung cancer and cardiovascular disease. *New England Journal of Medicine* 334:1150–1155.

Omland, T., J. A. de Lemos, M. S. Sabatine, C. A. Christophi, M. M. Rice, K. A. Jablonski, S. Tjora, M. J. Domanski, B. J. Gersh, J. L. Rouleau, M. A. Pfeffer, E. Braunwald, and the Prevention of Events with Angiotensin Converting Enzyme Inhibition (PEACE) Trial Investigators. 2009. A sensitive cardiac troponin T assay in stable coronary artery disease. *New England Journal of Medicine* 361(26):2538–2547.

Oram, J. F. 2001. Novel approaches to treating cardiovascular disease: Lessons from Tangier disease. *Expert Opinion on Investigational Drugs* 10(3):427–438.

Packard, R. R., and P. Libby. 2008. Inflammation in atherosclerosis: From vascular biology to biomarker discovery and risk prediction. *Clinical Chemistry* 54(1):24–38.

Paffen, E., and M. P. DeMaat. 2006. C-reactive protein in atherosclerosis: A causal factor? *Cardiovascular Research* 71(1):30–39.

Pai, V. B., and M. C. Nahata. 2000. Cardiotoxicity of chemotherapeutic agents. Incidence, treatment and prevention. *Drug Safety* 22(4):263–302.

Pantaleo, M. A., M. Nannini, E. Lopci, P. Castellucci, A. Maleddu, F. Lodi, C. Nanni, V. Allegri, M. Astorino, G. Brandi, M. Di Battista, S. Boschi, S. Fanti, and G. Biasco. 2008a. Molecular imaging and targeted therapies in oncology: New concepts in treatment response assessment. A collection of cases. *International Journal of Oncology* 33(3):443–452.

Pantaleo, M. A., M. Nannini, A. Maleddu, S. Fanti, V. Ambrosini, C. Nanni, S. Boschi, and G. Biasco. 2008b. Conventional and novel PET tracers for imaging in oncology in the era of molecular therapy. *Cancer Treatment Reviews* 34(2):103–121.

Panteghini, M., W. Gerhardt, F. S. Apple, F. Dati, J. Ravkilde, and A. H. Wu. 2001. Quality specifications for cardiac troponin assays. *Clinical Chemistry and Laboratory Medicine* 39(2):175–179.

Park, J. O., S. I. Lee, S. Y. Song, K. Kim, W. S. Kim, C. W. Jung, Y. S. Park, Y. H. Im, W. K. Kang, M. H. Lee, K. S. Lee, and K. Park. 2003. Measuring response in solid tumors: Comparison of RECIST and WHO response criteria. *Japanese Journal of Clinical Oncology* 33(10):533–537.

Pasceri, V., J. T. Willerson, and E. T. H. Yeh. 2000. Direct proinflammatory effect of C-reactive protein on human endothelial cells. *Circulation* 102(18):2165–2168.

Paul, A., K. W. Ko, L. Li, V. Yechoor, M. A. McCrory, A. J. Szalai, and L. Chan. 2004. C-reactive protein accelerates the progression of atherosclerosis in apolipoprotein E-deficient mice. *Circulation* 109(5):647–655.

Pearson, T. 2007. The prevention of cardiovascular disease: Have we really made progress? *Health Affairs* 26(1):49–60.

Pearson, T. A., G. A. Mensah, R. W. Alexander, J. L. Anderson, R. O. Cannon, M. Criqui, Y. Y. Fadl, S. P. Fortmann, Y. Hong, G. L. Myers, N. Rifai, S. C. Smith, K. Taubert, R. P. Tracy, and F. Vinicor. 2003. Markers of inflammation and cardiovascular disease: Application to clinical and public health practice. *Circulation* 107(3):499–511.

Pederson, K. O. 1947. On a low-density lipoprotein appearing in normal human plasma. *Journal of Physical Chemistry* 51:156–163.

Pepys, M. B. 2005. CRP or not CRP? That is the question. *Arteriosclerosis, Thrombosis, and Vascular Biology* 25(6):1091–1094.

Pepys, M. B., I. F. Rowe, and M. L. Baltz. 1985. C-reactive protein: Binding to lipids and lipoproteins. *International Review of Experimental Pathology* 27:83–111.

Pepys, M. B., G. M. Hirschfield, G. A. Tennent, J. R. Gallimore, M. C. Kahan, V. Bellotti, P. N. Hawkins, R. M. Myers, M. D. Smith, A. Polara, A. J. A. Cobb, S. V. Ley, J. A. Aquilina, C. V. Robinson, I. Sharif, G. A. Gray, C. A. Sabin, M. C. Jenvey, S. E. Kolstoe, D. Thompson, and S. P. Wood. 2006. Targeting C-reactive protein for the treatment of cardiovascular disease. *Nature* 440(27):1217–1221.

Pineda, J., F. Marin, P. Marco, V. Roldan, J. Valencia, J. M. Ruiz-Nodar, D. H. Romero, F. Sogorb, and G. Y. Lip. 2009. The prognostic value of biomarkers after a premature myocardial infarction. *International Journal of Cardiology* EPub ahead of print.

Plump, A. S., J. D. Smith, T. Hayek, K. Aalto-Setala, A. Walsh, J. G. Verstuyft, E. M. Rubin, and J. L. Breslow. 1992. Severe hypercholesterolemia and atherosclerosis in apolipoprotein E- deficient mice created by homologous recombination in ES cells. *Cell* 71(2):343–353.

Poole, S., D. Walker, R. E. Gaines Das, J. R. Gallimore, and M. B. Pepys. 1998. The first international standard for serum amyloid A protein (SAA). Evaluation in an international collaborative study. *Journal of Immunological Methods* 214(1–2):1–10.

Pugnale, N., F. Waridel, H. Bouzourène, A. Boubaker, M. Pugnale, R. C. Gaillard, and F. Gomez. 2003. Pharyngeal pituitary non-functioning adenoma with normal intra-sellar gland: Massive tumor shrinkage on octreotide therapy. *European Journal of Endocrinology* 148(3):357–364.

Rader, D. J. 2006. Molecular regulation of HDL metabolism and function: Implications for novel therapies. *Journal of Clinical Investigation* 116(12):3090–3100.

Ratain, M. J., R. Mick, R. L. Schilsky, and M. Siegler. 1993. Statistical and ethical issues in the design and conduct of phase I and II clinical trials of new anticancer agents. *Journal of the National Cancer Institute* 85(20):1637–1643.

Ridker, P. M. 2007. C-reactive protein and the prediction of cardiovascular events among those at intermediate risk. *Journal of the American College of Cardiology* 49(21):2129–2138.

Ridker, P. M., M. Cushman, M. J. Stampfer, R. P. Tracy, and C. H. Hennekens. 1997. Inflammation, aspirin, and the risk of cardiovascular disease in apparently healthy men. *The New England Journal of Medicine* 336(14):973–979.

Ridker, P. M., N. Rifai, M. A. Pfeffer, F. M. Sacks, L. A. Moye, S. Goldman, G. C. Flaker, and E. Braunwald. 1998. Inflammation, pravastatin, and the risk of coronary events after myocardial infarction in patients with average cholesterol levels. Cholesterol Recurrent Events (CARE) Investigators. *Circulation* 98(9):839–844.

Ridker, P. M., N. Rifai, M. Clearfield, J. R. Downs, S. E. Weis, J. S. Miles, and A. M. Gotto, Jr. 2001. Measurement of c-reactive protein for the targeting of statin therapy in the primary prevention of acute coronary events. *New England Journal Medicine* 344(26): 1959-1965).

Ridker, P. M., E. Danielson, F. A. H. Fonseca, J. Genest, A. M. Gotto Jr., J. J. P. Kastelein, W. Koenig, P. Libby, A. J. Lorenzatti, J. G. MacFadyen, B. G. Nordestgaard, J. Shepherd, J. T. Willerson, and R. J. Glynn. 2008a. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. *New England Journal of Medicine* 359(21):2195–2207.

Ridker, P. M., N. P. Paynter, N. Rifai, J. M. Caziano, and N. R. Cook. 2008b. C-reactive protein and parental history improve global cardiovascular risk prediction. The Reynolds Risk Score for Men. *Circulation* 118(22):2243–2251.

Ridker, P. M., E. Danielson, F. A. H. Fonseca, J. Genest, A. M. Gotto, Jr., J. J. P. Kastelein, W. Koenig, P. Libby, A. J. Lorenzatti, J. G. MacFadyen, B. G. Nordestgaard, J. Shepherd, J. T. Willerson, R. J. Glynn, and JUPITER Trial Study Group. 2009. Reduction in C-reactive protein and LDL cholesterol and cardiovascular event rates after initiation of rosuvastatin: A prospective study of the JUPITER trial. *The Lancet* 373(9670):1175–1182.

Roberts, W. L. 2004. CDC/AHA Workshop on Markers of Inflammation and Cardiovascular Disease: Application to Clinical and Public Health Practice: Laboratory tests available to assess inflammation—Performance and standardization: A background paper. *Circulation* 110(25):e572–e576.

Roberts, W. L., C. B. Calcote, B. K. De, V. Holmstrom, C. Narlock, F. S. Apple, and D. Whitehair. 1997. Prevention of analytical false-positive increases of cardiac troponin I on the Stratus II analyzer. *Clinical Chemistry* 43(5):860–861.

Robins, S. J. 2001. Targeting low high-density lipoprotein cholesterol for therapy: Lessons from the Veterans Affairs High-density Lipoprotein Intervention Trial. *American Journal of Cardiology* 88(12A):19N–23N.

Ross, R. 1993. The pathogenesis of atherosclerosis: A perspective for the 1990s. *Nature* 362(6423):801–809.

Ross, R. 1999. Atherosclerosis—An inflammatory disease. *The New England Journal of Medicine* 340(2):115–126.

Rubins, H. B., S. J. Robins, D. Collins, C. L. Fye, J. W. Anderson, M. B. Elam, F. H. Faas, E. Linares, E. J. Schaefer, G. Schectman, T. J. Wilt, and J. Wittes. 1999. Gemfibrozil for the secondary prevention of coronary heart disease in men with low levels of high-density lipoprotein cholesterol. Veterans Affairs High-Density Lipoprotein Cholesterol Intervention Trial Study Group. *New England Journal of Medicine* 341(6):410–418.

Sacks, F. M., and H. Campos. 2003. Clinical review 163: Cardiovascular endocrinology: Low-density lipoprotein size and cardiovascular disease: A reappraisal. *Journal of Clinical Endocrinology and Metabolism* 88(10):4525–4532.

Sakatini, T., T. Shirayama, Y. Suzaki, T. Yamamoto, H. Mani, T. Kawasaki, H. Sugihara, and H. Matsubara. 2005. The association between cholesterol and mortality in heart failure. Comparison between patients with and without coronary artery disease. *International Heart Journal* 46(4):619–629.

Sandri, M. T., D. Cardinale, L. Zorzino, R. Passerini, P. Lentati, A. Martinoni, G. Martinelli, and C. M. Cipolla. 2003. Minor increases in plasma troponin I predict decreased left ventricular ejection fraction after high-dose chemotherapy. *Clinical Chemistry* 49(2):248–252.

Sargent, D. J., L. Rubinstein, L. Schwartz, J. E. Dancey, C. Gatsonis, L. E. Dodd, and L. K. Shankar. 2009. Validation of novel imaging methodologies for use as cancer clinical trial end-points. *European Journal of Cancer* 45(2):290–299.

Sattar, N., D. Preiss, H. M. Murray, P. Welsh, B. M. Buckley, A. J. de Craen, S. R. Seshasai, J. J. McMurray, D. J. Freeman, J. W. Jukema, P. W. Macfarlane, C. J. Packard, D. J. Stott, R. G. Westendorp, J. Shepherd, B. R. Davis, S. L. Pressel, R. Marchioli, R. M. Marfisi, A. P. Maggioni, L. Travazzi, G. Tognoni, J. Kjekshus, T. R. Pedersen, T. J. Cook, A. M. Gotto, M. B. Clearfield, J. R. Downs, H. Nakamura, Y. Ohashi, K. Mizuno, K. K. Ray, I. Ford. 2010. Statins and risk of incident diabetes: A collaborative meta-analysis of randomised statin trials. *Lancet* 375(9716):735–742.

Scanu, A. 1966. Forms of human serum high density lipoprotein protein. *Journal of Lipid Research* 7(2):295–306.

Schatz, I. J., K. Masaki, K. Yano, R. Chen, B. L. Rodriguez, and J. D. Curb. 2001. Cholesterol and all-cause mortality in elderly people from the Honolulu Heart Program: A cohort study. *Lancet* 358(9279):351–355.

Schectman, G., M. Patsches, and E. A. Sasse. 1996. Variability in cholesterol measurements: Comparison of calculated and direct LDL cholesterol determinations. *Clinical Chemistry* 42(5):732–737.

Schepkin, V. D., T. L. Chenevert, K. Kuszpit, K. C. Lee, C. R. Meyer, T. D. Johnson, A. Rehemtulla, and B. D. Ross. 2006. Sodium and proton diffusion MRI as biomarkers for early therapeutic response in subcutaneous tumors. *Magnetic Resonance Imaging* 24(3):273–278.

Scher, H. I., S. Halabi, I. Tannock, M. Morris, C. N. Sternberg, M. A. Carducci, M. A. Eisenberger, C. Higano, G. J. Bubley, R. Dreicer, D. Petrylak, P. Kantoff, E. Basch, W. K. Kelly, W. D. Figg, E. J. Small, T. M. Beer, G. Wilding, A. Martin, and M. Hussain. 2008. Design and end points of clinical trials for patients with progressive prostate cancer and castrate levels of testosterone: Recommendations of the Prostate Cancer Clinical Trials Working Group. *Journal of Clinical Oncology* 26(7):1148–1159.

Schwartz, G. G., A. G. Olsson, M. D. Ezekowitz, P. Ganz, M. F. Oliver, D. Waters, A. Zeiher, B. R. Chaitman, S. Leslie, and T. Stern. 2001. Effects of atorvastatin on early recurrent ischemic events in acute coronary syndromes. The MIRACL study: A randomized controlled trial. *Journal of the American Medical Association* 285(13):1711–1718.

Scirica, B. M., and D. A. Morrow. 2004. Troponins in acute coronary syndromes. *Progress in Cardiovascular Diseases* 47(3):177–188.

Seibert, R. M., C. R. Ramsey, J. W. Hines, P. A. Kupelian, K. M. Langen, S. L. Meeks, and D. D. Scaperoth. 2007. A model for predicting lung cancer response to therapy. *International Journal of Radiation Oncology, Biology, Physics* 67(2):601–609.

Sevinc, A., and N. S. Turhal. 2008. Please, desist RECIST criteria in GIST, at least in me. *Onkologie* 31(10):556.

Shah, S. H., and J. A. deLemos. 2009. Biomarkers and cardiovascular disease: Determining causality and quantifying contribution to risk assessment. *Journal of the American Medical Association* 301(1):92–93.

Shah, V. P., J. P. Casas, J. A. Cooper, I. Tzoulaki, R. Sofat, V. McCormack, L. Smeeth, J. E. Deanfield, G. D. Lowe, A. Rumley, F. G. R. Fowkes, S. E. Humphries, and A. D. Hingorani. 2009. Critical appraisal of CRP measurement for the prediction of coronary heart disease events: New data and systematic review of 31 prospective cohorts. *International Journal of Epidemiology* 38(1):217–231.

Sharma, U., K. K. Danishad, V. Seenu, and N. R. Jagannathan. 2009. Longitudinal study of the assessment by MRI and diffusion-weighted imaging of tumor response in patients with locally advanced breast cancer undergoing neoadjuvant chemotherapy. *NMR in Biomedicine* 22(1):104–113.

Shepherd, J., S. M. Cobbe, I. Ford, C. G. Isles, A. R. Lorimer, P. W. MacFarlane, J. H. McKillop, and C. J. Packard. 1995. Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. *New England Journal of Medicine* 333(20):1301–1307.

Shepherd, J., G. J. Blauw, M. B. Murphy, E. L. Bollen, B. M. Buckley, S. M. Cobbe, I. Ford, A. Gaw, M. Hyland, J. W. Jukema, A. M. Kamper, P. W. MacFarlane, A. E. Meinders, J. Norrie, C. J. Packard, I. J. Perry, D. J. Stott, B. J. Sweeney, C. Twomey, and R. G. Westendorp. 2002. Pravastatin in elderly individuals at risk of vascular disease (PROSPER): A randomized controlled trial. *Lancet* 360:1623–1630.

Singh, U., S. Devaraj, and I. Jialal. 2005. C-reactive protein decreases tissue plasminogen activator activity in human aortic endothelial cells: Evidence that C-reactive protein is a procoagulant. *Arteriosclerosis, Thrombosis, and Vascular Biology* 25(10):2216–2221.

Sniderman, A. D. 2002. How, when, and why to use apolipoprotein B in clinical practice. *American Journal of Cardiology* 90(8A):48i–54i.

Sniderman, A. D., and S. M. Marcovina. 2006. Apolipoprotein A1 and B. *Clinics in Laboratory Medicine* 26(4):733–750.

Specht, J. M., S. L. Tam, B. F. Kurland, J. R. Gralow, R. B. Livingston, H. M. Linden, G. K. Ellis, E. K. Schubert, L. K. Dunnwald, and D. A. Mankoff. 2007. Serial 2-[18F] fluoro-2-deoxy-D-glucose positron emission tomography (FDG-PET) to monitor treatment of bone-dominant metastatic breast cancer predicts time to progression (TTP). *Breast Cancer Research and Treatment* 105(1):87–94.

Spies, C., V. Haude, R. Fitzner, K. Schröder, M. Overbeck, N. Runkel, and W. Schaffartzik. 1998. Serum cardiac troponin T as a prognostic marker in early sepsis. *Chest* 113(4):1055–1063.

St-Pierre, A. C., B. Cantin, G. R. Dagenais, P. Mauriege, P. M. Bernard, J. P. Despres, and B. Lamarche. 2005. Low-density lipoprotein subfractions and the long-term risk of ischemic heart disease in men: 13-year follow-up data from the Quebec Cardiovascular Study. *Arteriosclerosis Thrombosis and Vascular Biology* 25(3):553–559.

Stacchiotti, S., P. Collini, A. Messina, C. Morosi, M. Barisella, R. Bertulli, C. Piovesan, P. Dileo, V. Torri, A. Gronchi, and P. G. Casali. 2009. High-grade soft-tissue sarcomas: Tumor response assessment—Pilot study to assess the correlation between radiologic and pathologic response by using RECIST and Choi criteria. *Radiology* 251(2):447–456.

Stadler, W. M., and M. J. Ratain. 2000. Development of target-based antineoplastic agents. *Investigational New Drugs* 18(1):7–16.

Stamler, J., O. Vaccaro, J. D. Neaton, and D. Wentworth. 1993. Diabetes, other risk factors, and 12-yr cardiovascular mortality for men screened in the Multiple Risk Factor Intervention Trial. *Diabetes Care* 16(2):434–444.

Stegger, L., M. Schäfers, M. Weckesser, and O. Schober. 2008. EANM-ESR white paper on multimodality imaging. *European Journal of Nuclear Medicine and Molecular Imaging* 35(3):677–680.

Steinberg, D. 1997. A critical look at the evidence for the oxidation of LDL in atherogenesis. *Atherosclerosis* 131(Suppl):S5–S7.

Steinberg, D., S. Parthasarathy, T. E. Carew, J. C. Khoo, and J. L. Witztum. 1989. Beyond cholesterol: Modifications of low-density lipoprotein that increase its atherogenicity. *New England Journal of Medicine* 320(14):915–924.

Stewart, D. J. 2008. Re: Design of phase II cancer trials using a continuous endpoint of change in tumor size: Application to a study of sorafenib and erlotinib in non-small cell lung cancer. *Journal of the National Cancer Institute* 100(6):445–446; author reply 446.

Strassburg, J., T. Junginger, T. Trinh, O. Püttcher, K. Oberholzer, R. J. Heald, and P. Hermanek. 2008. Magnetic resonance imaging (MRI)-based indication for neoadjuvant treatment of rectal carcinoma and the surrogate endpoint CRM status. *International Journal of Colorectal Disease* 23(11):1099–1107.

Tall, A. R. 1993. Plasma cholesteryl ester transfer protein. *Journal of Lipid Research* 34(8):1255–1274.

Tardif, J.-C., T. Heinonen, D. Orloff, and P. Libby. 2006. Vascular biomarkers and surrogates in cardiovascular disease. *Circulation* 113(25):2936–2942.

Tardif, J. C., J. Gregoire, P. L. L'Allier, R. Ibrahim, J. Lesperance, T. M. Heinonen, S. Kouz, C. Berry, R. Basser, M. A. Lavoie, M. C. Guertin, and J. Rodes-Cabau. 2007. Effects of reconstituted high-density lipoprotein infusions on coronary atherosclerosis: A randomized controlled trial. *Journal of the American Medical Association* 297(15):1675–1682.

Taylor, K. E., J. C. Giddings, and C. W. van den Berg. 2005. C-reactive protein-induced in vitro endothelial cell activation is an artefact caused by azide and lipopolysaccharide. *Arteriosclerosis, Thrombosis, and Vascular Biology* 25(6):1225–1230.

Therasse, P., E. A. Eisenhauer, and J. Verweij. 2006. RECIST revisited: A review of validation studies on tumour assessment. *European Journal of Cancer* 42(8):1031–1039.

Thom, T., N. Haase, W. Rosamond, V. Howard, J. Rumsfeld, T. Manolio, Z. Zheng, K. Flegal, C. O'Donnell, S. Kittner, D. Lloyd-Jones, D. J. Goff, Y. Hong, R. Adams, G. Friday, K. Furie, P. Gorelick, B. Kissela, J. Marler, J. Meigs, V. Roger, S. Sidney, P. Sorlie, J. Steinberger, S. Wasserthiel-Smoller, M. Wilson, and P. Wolf. 2006. Heart disease and stroke statistics—2006 update: A report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation* 113(6):e85–e151.

Thygesen, K., J. S. Alpert, H. D. White, Biomarker Group, A. S. Jaffe, F. S. Apple, M. Galvani, H. A. Katus, L. K. Newby, J. Ravkilde, ECG Group, B. Chaitman, P. M. Clemmensen, M. Dellborg, H. Hod, P. Porela, Imaging Group, R. Underwood, J. J. Bax, G. A. Beller, R. Bonow, E. E. Van Der Wall, Intervention Group, J.-P. Bassand, W. Wijns, T. B. Ferguson, P. G. Steg, B. F. Uretsky, D. O. Williams, Clinical Investigation Group, P. W. Armstrong, E. M. Antman, K. A. Fox, C. W. Hamm, E. M. Ohman, M. L. Simoons, Global Perspective Group, P. A. Poole-Wilson, E. P. Gurfinkel, J.-L. Lopez-Sendon, P. Pais, S. Mendis, J.-R. Zhu, Implementation Group, L. C. Wallentin, F. Fernandez-Aviles, K. M. Fox, A. N. Parkhomenko, S. G. Priori, M. Tendera, L.-M. Voipio-Pulkki, ESC Committee for Practice Guidelines, A. Vahanian, A. J. Camm, R. De Caterina, V. Dean, K. Dickstein, G. Filippatos, C. Funck-Brentano, I. Hellemans, S. D. Kristensen, K. McGregor, U. Sechtem, S. Silber, P. Widimsky, J. L. Zamorano, Document Reviewers, J. Morais, S. Brener, R. Harrington, D. Morrow, M. Lim, M. A. Martinez-Rios, S. Steinhubl, G. N. Levine, W. B. Gibler, D. Goff, M. Tubaro, D. Dudek, and N. Al-Attar. 2007. Universal definition of myocardial infarction: Kristian Thygesen, Joseph S. Alpert and Harvey D. White on behalf of the Joint ESC/ACCF/AHA/WHF Task Force for the Redefinition of Myocardial Infarction. *Eur Heart J* 28(20):2525–2538.

Tran, L. N., M. S. Brown, J. G. Goldin, X. Yan, R. C. Pais, M. F. McNitt-Gray, D. Gjertson, S. R. Rogers, and D. R. Aberle. 2004. Comparison of treatment response classifications between unidimensional, bidimensional, and volumetric measurements of metastatic lung lesions on chest computed tomography. *Academic Radiology* 11(12):1355–1360.

Tuma, R. S. 2006. Sometimes size doesn't matter: Reevaluating RECIST and tumor response rate endpoints. *Journal of the National Cancer Institute* 98(18):1272–1274.

Ullrich, R. T., T. Zander, B. Neumaier, M. Koker, T. Shimamura, Y. Waerzeggers, C. L. Borgman, S. Tawadros, H. Li, M. L. Sos, H. Backes, G. I. Shapiro, J. Wolf, A. H. Jacobs, R. K. Thomas, and A. Winkeler. 2008. Early detection of erlotinib treatment response in NSCLC by 3'-deoxy-3'-[F]-fluoro-L-thymidine ([F]FLT) positron emission tomography (PET). *PLoS One* 3(12):e3908.

van Aalst-Cohen, E. S., A. C. Jansen, S. de Jongh, P. R. de Sauvage Nolting, and J. J. Kastelein. 2004. Clinical, diagnostic, and therapeutic aspects of familial hypercholesterolemia. *Seminars in Vascular Medicine* 4(1):31–41.

Van Beers, B. E., and V. Vilgrain. 2008. Biomarkers in abdominal imaging. *Abdominal Imaging* 34(6):663–667.

Van den Berg, C. W., and K. E. Taylor. 2005. Letter in response to Bisoendial et al.: "Activation of inflammation and coagulation after infusion of C-reactive protein in humans." *Circulation Research* 97(1):e2.

Verdecchia, P., G. Reboldi, F. Angeli, R. Gattobigio, C. Borgioni, L. Filippucci, F. Poeta, and C. Porcellati. 2004. Prognositc value of lipoprotein fractions in essential hypertension. *Blood Pressure* 13(5):295–303.

Verschuren, W. M., D. R. Jacobs, B. P. Bloemberg, D. Kromhout, A. Menotti, C. Aravanis, H. Blackburn, R. Buzina, A. S. Dontas, F. Fidanza, M. J. Karvonen, S. Nedeljkovic, A. Nissinen, and H. Toshima. 1995. Serum total cholesterol and long-term coronary heart disease mortality in different cultures: Twenty-five-year follow-up of the seven countries study. *Journal of the American Medical Association* 274(2):131–136.

Virella, G., and M. F. Lopes-Virella. 2008. Atherogenesis and the humoral immune response to modified lipoproteins. *Atherosclerosis* 200(2):239–246.

Vokurka, E. A., A. Herwadkar, N. A. Thacker, R. T. Ramsden, and A. Jackson. 2002. Using Bayesian tissue classification to improve the accuracy of vestibular schwannoma volume and growth measurement. *American Journal of Neuroradiology* 23(3):459-467.

Wahl, R. L., H. Jacene, Y. Kasamon, and M. A. Lodge. 2009. From RECIST to PERCIST: Evolving considerations for PET Response Criteria in Solid Tumors. *Journal of Nuclear Medicine* 50(Suppl 1):122S–150S.

Wallace, T. W., S. M. Abdullah, M. H. Drazner, S. R. Das, A. Khera, D. K. McGuire, F. Wians, M. S. Sabatine, D. A. Morrow, and J. A. de Lemos. 2006. Prevalence and determinants of troponin T elevation in the general population. *Circulation* 113(16):1958–1965.

Walter, M. 2009. Interrelationships among HDL metabolism, aging, and atherosclerosis. *Arteriosclerosis, Thrombosis, and Vascular Biology* 29(9):1244–1250.

Wang, T. J., P. Gona, M. G. Larson, G. H. Tofler, D. Levy, C. Newton-Chen, P. F. Jacque, N. Rifai, J. Selhub, S. J. Robins, E. J. Benjamin, R. D'Agostino, and S. V. Ramachamdan. 2006. Multiple biomarkers for the prediction of first major cardiovascular events and death. *New England Journal of Medicine* 355(25):2631–2639.

Watanabe, Y., T. Ito, and M. Shiomi. 1985. The effect of selective breeding on the development of coronary atherosclerosis in WHHL rabbits: An animal model for familial hypercholesterolemia. *Atherosclerosis* 56:71–79.

Weber, W. A. 2009. Assessing tumor response to therapy. *Journal of Nuclear Medicine* 50(Suppl 1):1S–10S.

Welsh, P., C. J. Packard, and N. Sattar. 2008. Novel antecedent plasma biomarkers of cardiovascular disease: Improved evaluation methods and comparator benchmarks raise the bar. *Current Opinion in Lipidology* 19(6):563–571.

Weverling-Rijnsburger, A. W. E., G. J. Blauw, A. M. Lagaay, D. L. Knook, A. E. Meinders, and R. G. J. Westendorp. 1997. Total cholesterol and risk of mortality in the oldest old. *Lancet* 350(9085):1119–1123.

Whicher, J. T. 1998. BCR/IFCC reference material for plasma proteins (CRM 470). Community Bureau of Reference. International Federation of Clinical Chemistry. *Clinical Biochemistry* 31(6):459–465.

Whitton, C. M., D. Sands, A. R. Hubbard, and P. J. Gaffney. 2000. A collaborative study to establish the 2nd International Standard for Fibrinogen, Plasma. *Thrombosis and Haemostasis* 84(2):258–262.

Williams, P. T., H. R. Superko, W. L. Haskell, E. L. Alderman, P. J. Blanche, L. G. Holl, and R. M. Krauss. 2003. Smallest LDL particles are most strongly related to coronary disease progression in men. *Arteriosclerosis, Thrombosis, and Vascular Biology* 23(2):314–321.

Wood, L. G., D. A. Fitzgerald, P. G. Gibson, D. M. Cooper, C. E. Collins, and M. L. Garg. 2001. Oxidative stress in cystic fibrosis: Dietary and metabolic factors. *Journal of the American College of Nutrition* 20(2):157–165.

Wu, A. H. B., and A. S. Jaffe. 2008. The clinical need for high-sensitivity cardiac troponin assays for acute coronary syndromes and the role for serial testing. *American Heart Journal* 155(2):208–214.

Yamada, S., T. Gotoh, Y. Nakashima, K. Kayaba, S. Ishikawa, N. Nago, Y. Nakamura, Y. Itoh, and E. Kajii. 2001. Distribution of serum C-reactive protein and its association with atherosclerotic risk factors in a Japanese population: Jichi Medical School Cohort Study. *American Journal of Epidemiology* 153(12):1183–1190.

Yang, G. C., J. D. Goldberg, and P. X. Ye. 2003. Risk of malignancy in follicular neoplasms without nuclear atypia: Statistical analysis of 397 thyroidectomies. *Endocrine Practice* 9(6):510-516.

Yaron, G., A. Brill, O. Dashevsky, I. M. Yosef-Levi, E. Grad, H. D. Danenberg, and D. Varon. 2006. C-reactive protein promotes platelet adhesion to endothelial cells: A potential pathway in atherothrombosis. *British Journal of Haematology* 134(4):426–431.

Young, H., R. Baum, U. Cremerius, K. Herholz, O. Hoekstra, A. A. Lammertsma, J. Pruim, and P. Price. 1999. Measurement of clinical and subclinical tumour response using [18F]-fluorodeoxyglucose and positron emission tomography: Review and 1999 EORTC recommendations. European Organization for Research and Treatment of Cancer (EORTC) PET Study Group. *European Journal of Cancer* 35(13):1773–1782.

Yu, H. H., R. Joubran, M. Asmi, T. Law, A. Spencer, M. Jouma, and N. Rifai. 2000. Agreement among four homocysteine assays and results in patients with coronary atherosclerosis and controls. *Clinical Chemistry* 46(2):258–264.

Zacho, J., A. Tybjaerg-Hansen, J. S. Jensen, P. Grande, H. Sillesen, and B. G. Nordestgaard. 2008. Genetically elevated C-reactive protein and ischemic vascular disease. *New England Journal of Medicine* 359(1897–1908).

Zaninotto, M., M. M. Mion, E. Novello, S. Altinier, and M. Plebani. 2007. New biochemical markers: From bench to bedside. *Clinica Chimica Acta* 381(1):14–20.

Zethelius, B., N. Johnston, and P. Venge. 2006. Troponin I as a predictor of coronary heart disease and mortality in 70-year-old men: A community-based cohort study. *Circulation* 113(8):1071–1078.

Zhang, Y. X., W. K. Cliff, G. I. Schoefl, and G. Higgins. 1999. Coronary C-reactive protein distribution: Its relation to development of atherosclerosis. *Atherosclerosis* 145(2):375–379.

5

Strengthening Evidence-Based Regulation

CHAPTER RECOMMENDATIONS

The previous chapters introduced the committee's proposed biomarker evaluation framework and tested it using diverse case studies. While the committee's evaluation framework will provide a more complete review of the evidence supporting different contexts of biomarker use, the effective implementation of the framework may require additional actions by the Food and Drug Administration (FDA) and other stakeholders. First, the committee concluded that the FDA's current regulatory authority is not sufficient. Inadequate fulfillment of postmarketing studies and incomplete understanding of how consumers interpret food and dietary supplement claims prevent robust protection of public health. Therefore, the committee recommended that:

Recommendation 5:

5a. Congress should strengthen the FDA's authority to request and enforce postmarket surveillance across drugs, devices, and biologics when approvals are initially based on putative surrogate endpoint data.

5b. Congress should grant the FDA authority to request studies and sufficient authority to act on the results of studies on consumer understanding of claims on foods and supplements.

To support this recommendation, the first section of this chapter summarizes the FDA's role and outlines the FDA's regulatory authority and

describes some of the limitations related to the FDA's current regulatory capacities. Recommendation 5 has two parts due to the differing regulatory frameworks surrounding drugs, devices, and biologics as compared to foods and supplements. Its intent is parallel, nonetheless.

In addition to strengthened FDA regulatory capacity, the committee acknowledged that science-based decision making is reliant on the availability of scientific data. Although there are ongoing efforts to collect and analyze biomarker data, the committee concluded that these efforts are uneven and not optimally organized within the U.S. Department of Health and Human Services (HHS). Recognizing the value of a well-coordinated, comprehensive effort to collect and share biomarker information in advancing public health, the committee sought to improve ongoing biomarker data collection efforts. Improved FDA information infrastructure and surveillance systems may also enhance the agency's ability to interpret biomarkers and their relation to public health. Based on these findings, the committee made the following recommendation:

Recommendation 6:

6a. The U.S. Department of Health and Human Services should facilitate a coordinated, department-wide effort to encourage the collection and sharing of data about biomarkers for all uses, including drugs, biologics, devices, and foods.

6b. The FDA in coordination with other federal agencies should build needed data infrastructure and surveillance systems to handle the information necessary to gain sufficient understanding of the effects of biomarker use.

The second part of this chapter reviews the FDA's infrastructure capacity, and ongoing biomarker data collection efforts. Opportunities to facilitate data collection and sharing, such as precompetitive collaboration, will also be highlighted.

FDA REGULATORY AUTHORITY

Several federal agencies have responsibility for public health. In addition to the FDA and the other 10 agencies that comprise HHS, HHS also collaborates with units within the Departments of Defense, Veterans Affairs, Agriculture, and Education in carrying out its public health responsibilities. The FDA's mission is as follows:

> The FDA is responsible for protecting the public health by assuring the safety, efficacy, and security of human and veterinary drugs, biological products, medical devices, our nation's food supply, cosmetics, and products that emit radiation. The FDA is also responsible for advancing

> the public health by helping to speed innovations that make medicines and foods more effective, safer, and more affordable; and helping the public get the accurate, science-based information they need to use medicines and foods to improve their health. (FDA, 2008a)

The FDA's task is large. The FDA regulates products that comprise about 25 percent of consumer spending in the United States, which comes to more than $1 trillion in spending (Subcommittee on Science and Technology, 2007). The FDA's 2008 budget authority was $1.87 billion; with user fees added to this number, the FDA's total 2008 budget was $2.42 billion (Office of Budget, 2009). As stated by Wood (2008) in his article *Playing "Kick the FDA"—Risk-free to Players but Hazardous to Public Health*,

> Between 1988 and 2007, additional FDA responsibilities were imposed by 137 specific statutes, 18 statutes of general applicability, and 14 executive orders (Subcommittee on Science and Technology, 2007). At the same time, the FDA received a 2007 federal appropriation of only $1.57 billion—less than 75% of the budget for the school district in its home county in Maryland.

For another comparison, Coca-Cola's advertising budget in 2008 was $3 billion (Coca-Cola, 2009). The money spent to promote one company's products in one year is greater than the money spent to ensure the safety of products purchased with one out of every four consumer dollars in the United States.

Recommendation 7.1 from the Institute of Medicine (IOM) report *The Future of Drug Safety* stated that "to support improvements in drug safety and efficacy activities over a product's lifecycle, the committee recommends that the Administration should request and Congress should approve substantially increased resources in both funds and personnel for the Food and Drug Administration" (IOM, 2007b). Food safety is also a challenge because responsibilities are spread over multiple agencies (IOM, 1998). IOM reports on food safety have also pointed out the need for sufficient funding to support a science-based food safety system (IOM, 1998). The call for adequate resources to protect food and drug safety has also been sounded by the FDA's Science Board in its report *FDA Science and Mission at Risk* (Subcommittee on Science and Technology, 2007). The challenges facing the FDA as its duties expand and its resources shrink have also been noted by IOM committees and workshops (IOM, 2007a, 2007b) as well as other entities (GAO, 2009b; IOM, 2007a, 2007b; Wood, 2008).

With its large task and small budget, the FDA faces criticism from many directions: when there is an outbreak of illness caused by a foodborne pathogen, when there are pervasive safety problems in food plants,

when medical and food products imported from other countries are adulterated, when new life-saving drugs are not approved fast enough, and when unsafe drugs are taken by patients for years before their risks are recognized. Criticism comes from the public, industry, and government. Decisions from case law modify laws and regulations; these decisions do not always consider their impact beyond a particular case. Changing administrations and priorities within the executive branch also complicate the FDA's ability to be successful in protecting public health.

The FDA derives much of its regulatory authority from the Food, Drug, and Cosmetic Act (FDCA), which was originally passed in 1938 and has been amended over time (Box 5-1). From a relatively modest portfolio of activities in 1938, the responsibilities of the agency have continued to expand. In the past two decades, Congress has enacted more than 100 statutes that directly impact the FDA's regulatory responsibilities—an average of 6 statutes per year, in addition to its core objectives. All of these statutes require some type of FDA action, such as the development or implementation of regulations or guidance documents, or the establishment of new regulatory programs. Although the FDA's purview of responsibilities continue to expand, the FDA gained through appropriation only 646 employees, an increase of 9 percent, and lost more than $300 million to inflation (Subcommittee on Science and Technology, 2007). Figure 5-1 illustrates the scope of the FDA's regulatory responsibilities, as outlined in 2006. In terms of dollars, the Center for Devices and Radiological Health (CDRH) regulated manufacturers with industry sales of $110 billion; the Center for Food Safety and Applied Nutrition (CFSAN) regulated $417 billion worth of domestic food, $49 billion in imported food, $60 billion in cosmetics, and $18 billion in dietary supplements; and the Center for Drug Evaluation and Research (CDER) regulated $275 billion in pharmaceutical sales (Subcommittee on Science and Technology, 2007).

The most recent amendment to the FDCA was the Food and Drug Administration Amendments Act of 2007 (FDAAA). The law expanded FDA authority and reauthorized the Prescription Drug User Fee Act (PDUFA), the Medical Device User Fee and Modernization Act, the Best Pharmaceuticals for Children Act, and the Pediatric Research Equity Act (FDA, 2009c). FDAAA requires more FDA involvement in ensuring that clinical trials are incorporated into ClinicalTrials.gov and provides the FDA with additional requirements, authorities, and resources related to pre- and postmarket drug safety, including the ability to require postmarketing studies, clinical trials, safety labeling changes, and Risk Evaluation and Mitigation Strategies (REMS). In addition, FDAAA requires new reporting of adverse events related to food (FDA, 2009b).

The traditional tools of regulatory agencies include regulation, approval or disapproval of applications, and enforcement (Hamburg and

BOX 5-1
Expanding FDA Responsibilities

The modern regulatory functions of the Food and Drug Administration (FDA) began with the passage of the 1906 Pure Food and Drug Act, which prohibited interstate commerce in adulterated and misbranded food and drugs. The Food, Drug, and Cosmetic Act (FDCA), passed by Congress in 1938, overhauled the public health system by authorizing the FDA to require evidence of safety for new drugs, set standards for food, and conduct factory inspections.

Since 1938, the FDA's role has expanded enormously. Enactment of a series of statutes, beginning in the 1950s and continuing into the 1970s, provided the FDA with a much broader mandate. For example, the *Kefauver-Harris Amendments of 1962* strengthened rules for drug safety and required manufacturers to demonstrate effectiveness of drugs. In 1976, the FDCA was amended to apply safety and effectiveness safeguards to new medical devices. Significant amendments to the FDCA since 1980 are included below:

- Infant Formula Act of 1980 and as amended in 1986
- Drug Price Competition and Patent Term Restoration Act of 1984
- Nutrition Labeling and Education Act of 1990
- Safe Medical Devices Act of 1990
- Medical Device Amendments of 1992
- Prescription Drug User Fee Act of 1992
- Dietary Supplement Health and Education Act of 1994
- FDA Export Reform and Enhancement Act of 1996
- Food Quality Protection Act of 1996
- Animal Drug Availability Act of 1996
- The Food and Drug Administration Modernization Act of 1997
- Best Pharmaceuticals for Children Act of 2002
- Pediatric Research Equity Act of 2003
- The Minor Use and Minor Species Animal Health Act of 2004
- Dietary Supplement and Nonprescription Drug Consumer Protection Act of 2006
- Food and Drug Administration Amendments Act of 2007

SOURCES: FDA (2009e, 2009f); Subcommittee on Science and Technology (2007).

Sharfstein, 2009). Prior to the 1970s, FDA functions were primarily related to law enforcement (e.g., issues of adulteration and misbranding). Current regulatory responsibilities are science based, as most of the FDA's work has moved away from the court to regulatory decisions involving scientific competencies and technical knowledge. Across all centers, the core regulatory functions can be roughly divided into three categories: (1) premarket review, (2) marketed product adverse event surveillance and

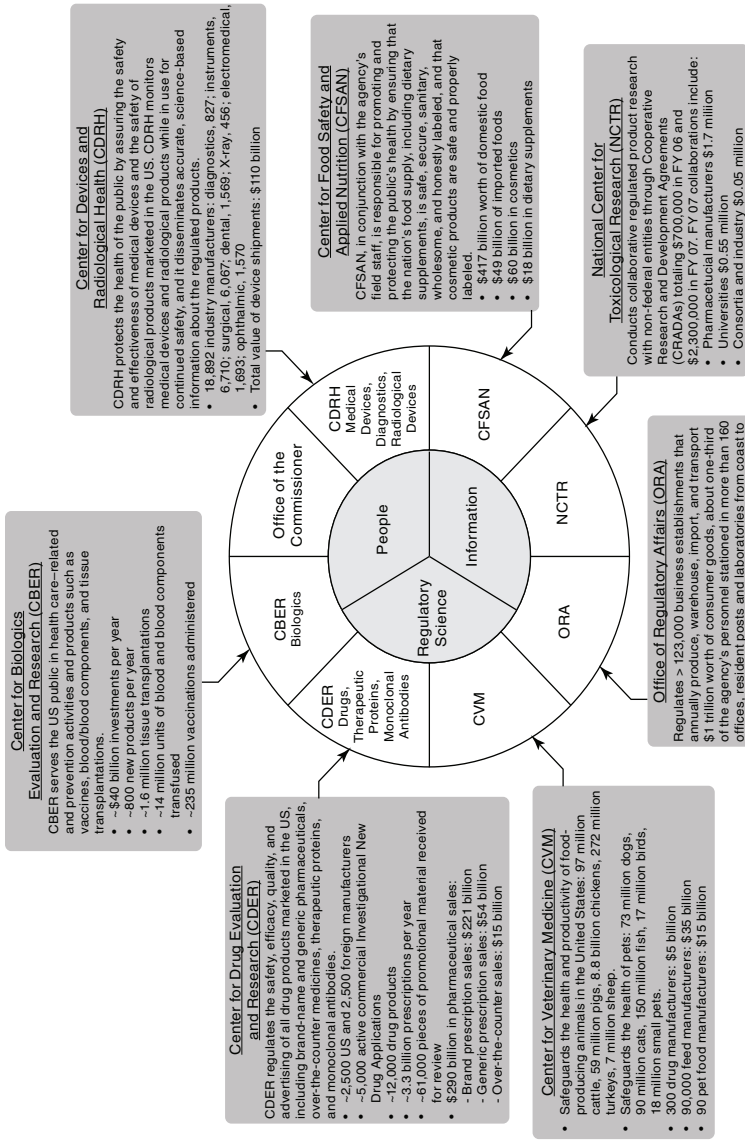

FIGURE 5-1 Food and Drug Administration—Regulatory Industry (FY2006): The people, science, and information needed to support innovation, grow industries, and protect the public both in our country and around the world.
SOURCE: Subcommittee on Science and Technology (2007).

BOX 5-2
Core FDA Regulatory Functions

- Premarket review involves reviewing and assessing sponsor applications and submissions and developing and/or implementing a Food and Drug Administration (FDA) response
  - Examples include Biologic License Applications, New Drug Applications, Pre-Market Notifications, Health Claim Petitions, Generally Recognized as Safe Notices; responses include approval, complete response, non-acceptance, non-approval, clearance, an information/advice letter, and other actions
- Marketed product adverse event surveillance and efficacy/safety assessment, which involves identifying or receiving information about marketed product safety or efficacy issues, assessing the information, and developing and/or implementing an FDA response
  - Examples include postmarket safety reports from sponsors or individuals; population-based surveillance studies; responses include inspections, follow-up and work with product sponsors, local public health agencies, recalls, label changes, issuance of health alerts, product recalls and withdrawals, and other legal action
- Ensuring marketed product quality and safety involve identifying or receiving information about a marketed product or a manufacturing or distribution facility, assessing this information, and developing or implementing an FDA response
  - Examples include information obtained from facility and product quality inspections; responses can include inspections, intensive follow-up and work with product sponsors or distributors, work with import agencies, product recalls and withdrawals, and legal action

SOURCE: Subcommittee on Science and Technology (2007).

efficacy/safety assessment, and (3) ensuring marketed product safety and quality (see Box 5-2) (Subcommittee on Science and Technology, 2007).

Drugs and Biologics

Two centers, CDER and the Center for Biologics Evaluation and Research (CBER), are responsible for assessing the safety and effectiveness of drugs and biologics, respectively. CDER is the largest of the FDA's five centers, and has responsibility for both prescription and over-the-counter drugs. CBER ensures the safety, purity, potency, and effectiveness of biological products, including vaccines, blood and blood products, cells, tissues, and gene therapies. The premarket review responsibilities of these

centers include reviewing and assessing sponsor applications and developing and/or implementing an FDA response for new, supplemental, or change-in-use products. Premarket applications related to these centers include Investigational New Drugs, New Drug Applications, and Biologic License Applications.

In addition to premarket review responsibilities, these centers are involved in adverse event surveillance and efficacy and safety assessment of marketed products. The FDA receives and analyzes reports of adverse events, participates in active surveillance and signal detection activities, takes action on safety and efficacy problems, and implements and evaluates risk communications. Particular concerns have been raised over the FDA's drug safety system, and a large component of the FDA's drug safety system consists of postmarketing surveillance activities. Previous reviews of the FDA, such as the IOM report *The Future of Drug Safety* (IOM, 2007b), suggest that the FDA has unclear and insufficient regulatory authorities related to enforcement. The following section outlines the limitations of current postmarking activities and the corresponding lack of FDA authority.

Postmarketing Surveillance in FDA Regulation of Drugs and Biologics

Phase I through III clinical trials are dedicated to demonstrating safety and effectiveness for FDA approval, and usually involve a few hundred to a few thousand individuals. However, many more individuals may ultimately receive the intervention post-FDA approval, and tracking clinical experience, through phase IV (postmarketing) studies is important for identifying relatively rare adverse events and determining effectiveness within different populations and circumstances. However, some evidence suggests that drug sponsors are not fulfilling their postmarketing obligations efficiently, and that the FDA lacks authority to hold drug sponsors to their commitments after drug approval.

Postmarketing surveillance may evaluate safety, efficacy, pharmacology, toxicology, and manufacturing controls, among other factors. The FDA requires drug applicants to conduct these studies in several situations. First, if a drug is approved under accelerated approval on the basis of a surrogate rather than clinical endpoint, then the FDA requires that postmarketing studies verify the safety and efficacy of the drug after it is on the market. If postmarketing studies do not substantiate clinical benefits, or raise safety concerns, the FDA may withdraw the drug. Second, in deferred pediatric studies, drugs approved in adults may be used in children with required postmarketing studies substantiating safety and efficacy in the pediatric populations. Third, when it is unethical to conduct clinical trials in humans, the FDA can approve drugs on

the basis of animal data, but also requires efficacy and safety data after approval. Finally, the FDA can request that the drug applicant conduct postmarketing studies prior to drug approval. Drug applicants agree to these commitments in writing, which the FDA then lists in its final drug approval letters.

The Food and Drug Administration Modernization Act of 1997 expanded the FDA's authority to oversee postmarketing study commitments. The legislation requires drug sponsors to report on the status of certain postmarketing studies (via annual status reports) and establishes that some information contained in these reports is considered public information. The FDA requires annual status reports for postmarketing studies that address clinical safety, clinical efficacy, clinical pharmacology, and nonclinical pharmacology.[1] In the annual status report, drug sponsors indicate the status of their postmarketing commitments, marking studies as pending (study has not started, but is not behind schedule), ongoing (ahead or on schedule), delayed, terminated (study ended before completion, but a final report has not been submitted to the FDA), or submitted.

There is concern that the current postmarketing surveillance system is inadequate in ensuring that drugs are safe and effective. The primary concern with the current system is that drug manufacturers are responsible for the collection, evaluation, and reporting of data from postmarketing studies of their own products. Statistics reveal that drug manufacturers are not efficiently fulfilling their postmarketing obligations; in 2004, fewer than half of promised postmarketing commitments had been initiated (Psaty et al., 2004). From 2004 to 2008, the number of open postmarketing commitments has remained relatively stable, at around 1,100–1,200, while the number of commitments met each year has also remained relatively stable, at around a much lower 100–160 each year.[2]

Drug manufacturers have little incentive to conduct timely postmarketing studies because these studies may reveal safety problems or other concerns that could result in more constrictive drug labeling, or withdrawal from the market, even though these studies may be a condition of drug approval. Drug manufacturers may also be tempted to conceal available data that suggest a drug has major risks. Examples of such concealment include unpublished data suggesting absence of benefit—or even risk of harm—of selective serotonin reuptake inhibitors in teens (Drummond, 2004), and data showing the interaction of cerivastatin with gemfibrozil and the risk of rhabdomyolysis. Although Bayer Corporation,

[1] 21 C.F.R. § 314.81(b)(2)(vii) (2003).

[2] 73 *Federal Register* 22157–22159 (2008), 72 *Federal Register* 5069–5070 (2007), 71 *Federal Register* 10978–10979 (2006), 70 Federal Register 8379–8381 (2005).

the manufacturer of cerivastatin, was aware of the risks of rhabdomyolysis as early as 4 months after the launch of the drug, the contraindication was not added to the package insert for more than 18 months[3] (Fontanarosa et al., 2004; Psaty et al., 2004).

Beyond the disincentives of drug manufacturers to reveal possible drug risks postapproval, there are also concerns that the FDA is ill equipped to oversee postmarketing activities and lacks regulatory recourse to ensure drug manufacturer compliance. In 2006, HHS conducted a review of the FDA's monitoring of its postmarketing study commitments, and came to two conclusions: first, the FDA cannot readily identify whether, or how timely, postmarketing study commitments are progressing toward completions; and second, that monitoring postmarketing study commitments are not a top priority at the FDA (Office of Inspector General, 2006). The report found that one-third of annual status reports were missing or incomplete and contained information with little utility, while the management information system supporting postmarketing activities was found to be ineffective. Compared to other priorities, postmarketing surveillance was also found to be of lower priority. Although PDUFA provided additional funding to the FDA to meet new time lines for drug approval, it prohibited the agency from using the user fees on postmarketing surveillance or other drug safety programs (Psaty et al., 2004), and FDA officials noted that PDUFA-associated activities (e.g., reviewing drug applications and documenting FDA/industry meetings) are of higher priority (Office of Inspector General, 2006). Furthermore, a Government Accountability Office (GAO) study found that there is a lack of criteria for determining which safety actions to take in response to postmarketing findings, and when they should occur (GAO, 2006).[4] In reviewing FDA postmarketing surveillance activities, the IOM report *The Future of Drug Safety* showed that the FDA lacks clear, unambiguous authority to enforce drug sponsor compliance with regulatory requirements. The committee recommended that Congress ensure that the FDA has the ability to require postmarketing risk assessment and risk management programs, and is equipped with better enforcement tools to ensure drug sponsor compliance (IOM, 2007b).

[3] Eventually Bayer voluntarily withdrew cerivastatin from the U.S. market in August 2001 due to the high rates of rhabdomyolysis.

[4] Also, a 2009 GAO analysis found that FDA has taken a passive approach to enforcing confirmatory study requirements and has never exercised its authority to withdraw a drug it approved based on surrogate endpoints under the accelerated approval process, even when such studies have been outstanding for nearly 13 years. This analysis found that two-thirds of postmarketing studies for drugs based on surrogate endpoints through the accelerated approval process have been closed from the creation of accelerated approval in 1992 to November 20, 2008 (GAO, 2009d).

In September 2007 FDAAA was enacted. Among its provisions were new authorities to require postmarket studies and clinical trials, safety labeling changes, and REMS. As a result of FDAAA, between March 25, 2008, and September 14, 2009, CDER and CBER issued 74 letters with postmarketing requirements to assess safety issues for drugs and biologics (FDA, 2009a). Whereas these kinds of studies would have had to have been undertaken voluntarily, they are now required with enforceable time lines. In addition, in 2008, the FDA introduced the Sentinel Initiative, a national integrated electronic database to detect adverse events of drugs and other medical products. It is hoped that the system will eventually monitor as many as 100 million individuals, and will be built from participating electronic health records and claims databases (IOM, 2009; Platt et al., 2009). The Sentinel System will be a distributed network, where all clinical data remains within the source systems' databases, with centralized software to query approved network questions.

Devices

The FDA Center for Devices and Radiological Health is responsible for regulating medical devices as well as radiation-emitting electronic products. CDRH categorizes devices into three classes, depending on this risks they pose. Class I devices, which include items such as tongue depressors, toothbrushes, and bedpans, have the lowest regulation and do not require review by the FDA prior to marketing. Class II devices face an intermediate level of regulation, including a clearance process that usually does not require submission of clinical data to the FDA. Class III devices, including implants and other high-risk devices, are the most regulated device category and require submission of clinical evidence of safety and effectiveness to secure FDA approval prior to marketing. Regulation of medical devices tends to lag behind the regulation of pharmaceuticals (IOM, 2005). In addition to establishing the three categories of devices, The Medical Device Amendments Act of 1976[5] also gave the FDA authority to create a system for reporting adverse events associated with devices. In 1984 the FDA issued regulations requiring manufacturers and importers of devices to report information indicating that a device might have caused or contributed to a death or serious injury.

The Safe Medical Devices Act of 1990 added requirements that hospitals and other facilities report to the FDA and manufacturers any events indicating that a device caused or contributed to an event. Additionally, the legislation established new requirements for manufacturers to track specific types of high-risk medical devices and gave the FDA authority

[5] Public Law 94-295.

to order recalls of devices in some circumstances. The 1990 legislation further provided that the FDA direct manufacturers to conduct additional information collection activities for certain implants and other devices with the potential to cause serious harm. While this Act increased the scope of device regulation, it also enabled certain medical devices for small user populations without requiring substantial clinical evidence of effectiveness through a Humanitarian Device Exemption.

The 1997 FDA Modernization Act reversed some provisions of the 1990 legislation, such as eliminating certain requirements for adverse event reporting, and ended provisions for mandatory postmarketing surveillance. The legislation prioritized FDA resources to higher risk devices and authorized a new adverse event reporting system based on a sample of hospitals and other user facilities. In addition to providing a system of user fees for FDA premarket reviews, the Medical Device User Fee and Modernization Act of 2002 authorized additional appropriations for postmarketing surveillance, but Congress did not appropriate the funds (reviewed by IOM, 2005).

Postmarketing Surveillance and Premarket Approval in FDA Regulation of Devices

Several factors indicate that the postmarketing surveillance activities for devices are inadequate. For the *Safe Medical Devices for Children* report (IOM, 2005), the IOM committee was asked to assess if the postmarket surveillance of medical devices provides adequate safeguards for pediatric populations. The committee found shortfalls in FDA performance that were, by and large, not limited to children. The committee found that the agency lacked effective procedures for monitoring the status of required postmarket studies, and there was a lack of public information regarding these studies. The committee recommended that the FDA establish a reliable system to track postmarket studies, and that this information should be publicly available (IOM, 2005). In 2009, the GAO added the FDA's oversight of medical products, including devices, to its list of high-risk areas warranting attention by Congress and the executive branch. In regard to postmarket surveillance of medical devices, the GAO reported that the number of adverse event reports associated with medical devices increased substantially from 2000 to 2006, but concluded that there are shortcomings in FDA's regulatory authority related to postmarket surveillance. According to the FDA, the volume of adverse event reports exceeds the agency's ability to consistently enter or review the reports in a routine manner. FDA officials told the GAO in 2008 that it had a number of strategies to prioritize the reviews of adverse event reports, but cannot review all of the reports received (GAO, 2009c).

In addition to limitations in postmarketing surveillance, the strength of evidence supporting premarketing device approvals has been questioned. Two recent reviews (Dhruva et al., 2009; Kramer et al., 2009) assessed the strength of studies used to support the premarket approval of cardiovascular devices. The first, sponsored by the FDA, found that more than 40 percent of studies used to approve cardiovascular devices lacked high quality data about either the treatment or safety goals of the study, and 25 percent of trials failed to adequately follow the outcomes of a sufficient number of patients (Kramer et al., 2009). The second study (Dhruva et al., 2009) found that 33 of 123 studies (27 percent) used to support the FDA approval of cardiovascular devices were randomized and 17 of 123 were blinded; 65 percent of premarket approvals were based on a single study. This review found that 187 of 213 (88 percent) primary endpoints were surrogate measures and 122 of 157 (78 percent) had a discrepancy between the number of patients enrolled in the study and the number analyzed. In recognition of concerns over the quality of studies used in device approval, the FDA is in the process of developing guidelines to set tougher scientific standards (Meier, 2009). Most likely, the FDA will urge device makers to use more sharply defined targets to measure success of clinical trials and more closely follow patients enrolled in trials to determine whether these targets are met.

Foods

Chapter 2 outlines the legal basis for health claims and describes the different types of claims (see Table 5-1). In addition, Chapter 2 discusses how the FDA has used biomarkers as surrogate endpoints in the evaluation of authorized and qualified health claims, and the current "surrogate endpoints of disease risk" used by CFSAN. The following section describes the FDA's regulatory authority related to foods, while Box 5-3 discusses FDA regulatory authority related to dietary supplements.[6]

Legislative mandates and legal action have influenced the way foods are regulated in the United States. Notably, the Nutrition Labeling and Education Act of 1990 (NLEA) first set the circumstances under which nutrition and health claims could be used. NLEA was designed to give consumers more scientifically valid information about foods they eat, and the statute directed the FDA to issue regulations providing for the use of

[6] Dietary supplements are defined as products that are intended to supplement the diet and contain one or more of the following "dietary ingredients": vitamins, minerals, herbs or other botanicals, amino acids, or other dietary supplements for use by humans to supplement the diet by increasing the total dietary intake, or concentrates, metabolites, constituents, extracts, or combinations for these ingredients.

TABLE 5-1 Types of Claims

Type of Claim	Description
Health claims based on significant scientific agreement (SSA), or authorized health claims	Health claims based on a high level of confidence in the validity of the relationship between the substance and the disease or health-related condition. Authorized health claim language must meet all regulatory requirements (e.g., development of hypertension or high blood pressure depends on many factors. [This product] can be part of a low-sodium, low-salt diet that might reduce the risk of hypertension or high blood pressure). Authorized health claims require Food and Drug Administration (FDA) review and approval.
Health claims based on authoritative statements	Claims based on authoritative statements about the substance/disease relationship by other scientific bodies of the U.S. government or the National Academy of Sciences.[a] Authoritative statements can be used without approval by an FDA review process.
Qualified health claims	Claims that do not meet the SSA standard can be allowed on the basis of lower evidence if the claim is accompanied by qualifying language. Qualified health claims require FDA review and approval.

statements that describe the relationship between a substance and disease in labeling of foods, including dietary supplements, after the statements have been reviewed and authorized by the FDA (CFSAN, 2009). Based on NLEA, the FDA authorizes health claims on the basis of significant scientific agreement (SSA), or:

> based on the totality of publicly available scientific evidence (including evidence from well-designed studies conducted in a manner which is consistent with generally recognized scientific procedures and principles), that there is significant scientific agreement among experts qualified by scientific training and experience to evaluate such claims, that the claim is supported by such evidence.[7]

[7] 21 C.F.R. § 101.14(c) (1998).

TABLE 5-1 Continued

Type of Claim	Description
Nutrient content claims	Claims that expressly or implicitly characterize the level of a nutrient (e.g., low in fat, high in vitamin C). The FDA accepts 11 core content claims, but only for substances with established Daily Reference Values or Reference Dietary Intakes.
Structure/function claims	Claims about the dietary impact of a nutrient on the structure or function of the human body, but cannot specify that the nutrient will cure, mitigate, treat, or prevent disease (e.g., calcium helps build strong bones). Does not require pre-review by the FDA.
Dietary guidance statements	Not considered a health claim, but appears on food labeling, and usually makes reference to a category of food rather than a specific substance (e.g., fruits and vegetables contribute to a healthful diet). Does not require pre-review by the FDA.

[a]In legislation, the term National Academy of Sciences refers to the whole of the National Academies.

SOURCES: CFSAN (2008); IFT (2005); Schneeman (2007); Taylor and Wilkening (2008); 21 C.F.R. § 101.74(e) (2009).

Legal action challenged the SSA standard and resulted in a process to allow claims with lesser scientific evidence, with qualifying language (qualified health claims) (Schneeman, 2007; Taylor and Wilkening, 2008).

Consumer Understanding and FDA Regulation in Food

In order to ensure the safety of nutrition-related uses of biomarkers, more research is needed on strategies for effective communication to consumers about health-related information. In the next paragraphs, research on communication for conveyance of health information related to foods is presented. Although the FDA has stipulated the numerous types of claims that manufacturers can include in product labeling, it can be difficult for consumers to assess the scientific merit of these

BOX 5-3
Dietary Supplements

Adding to the complexity of the FDA's regulatory authority, the FDA regulates dietary supplements under a different set of regulations than those covering conventional foods. Dietary supplements are defined in the Dietary Supplement Health and Education Act of 1994 (DSHEA) as food products, that among other things, are intended to supplement the diet and contain one or more of the following "dietary ingredients": vitamins, minerals, herbs or other botanicals, amino acids, or other dietary supplements for use by humans to supplement the diet by increasing the total dietary intake, or concentrates, metabolites, constituents, extracts, or combinations for these ingredients (FDA, 2009d). Prior to 1994, dietary supplements were subject to the same regulatory requirements as other foods. However, DSHEA created a new regulatory framework for the safety and some aspects of labeling of dietary supplements.

Under DSHEA, dietary supplements are broadly presumed to be safe, and the FDA does not have the authority to require dietary supplements to be approved for safety and efficacy before they enter the market (GAO, 2009a). Although the FDA must be notified of new dietary ingredients and the manufacturer must provide evidence that the new dietary ingredient is reasonably expected to be safe, the starting assumption is that dietary ingredients are safe (Yetley, 2007). The manufacturer is responsible for ensuring that any claims made about dietary supplements are substantiated by adequate evidence to show that they are not false or misleading (FDA, 2009d).

In contrast to authorized and qualified health claims for food, the FDA permits statements of nutritional support to be made in dietary supplement labeling without a scientific review of the evidence.[a] These claims are permitted under the following circumstances: (1) if the manufacturer has substantiation that a statement is truthful and not misleading; (2) if the labeling includes prominently displayed text that "This statement has not been evaluated by the Food and Drug Administration. This product is not intended to diagnose, treat, cure, or prevent any disease"; (3) the manufacturer notifies the FDA no later than 30 days after first marketing the dietary supplement with the statement.[b] In 2000, the FDA published final regulations on the types of unacceptable structure/function claims in the labeling of dietary supplements.[c]

Because of the regulatory framework for dietary supplements, the FDA has little information about the safety and effectiveness of dietary supplements. In 2007, the FDA issued a final rule establishing regulations to require good manufacturing practices to improve oversight of dietary supplements (FDA, 2007). However, a GAO report on dietary supplements reported that "consumers remain vulnerable to risks posed by potentially unsafe products" and found that the FDA's ability to

identify dietary supplement safety concerns is hindered by a lack of information (GAO, 2009a). With this limited information, a primary method for the FDA to identify safety concerns is postmarketing surveillance. In 2006, the Dietary Supplement and Nonprescription Drug Consumer Protection Act[d] amended the FDCA[e] to require dietary supplement companies that receive serious adverse event reports to submit information about the event to the FDA, effective December 22, 2007. Since the mandatory reporting requirement went into effect, the FDA has seen a three-fold increase in the number of total adverse events reported: from January through October 2008, the FDA received 948 adverse event reports, compared with 298 reports received over the same period in 2007. Of the 948 adverse event reports, 596 reports were mandatory reports of serious adverse events submitted by industry. However, the FDA recently estimated that the actual total number of adverse events (including mild, moderate, and serious adverse events) related to dietary supplements is more than 50,000, suggesting underreporting, limiting the amount of information the FDA receives (GAO, 2009a).

Once the FDA has identified a potential safety problem, the agency has several options, including issuing warning letters and consumer alerts, working with a company on a voluntary product recall, and banning an ingredient, among other options. However, the FDA's authority to respond is limited. According to the recent GAO report, the limitations are two-fold: first, the FDA dedicates relatively few resources to dietary supplement oversight activities; and second, under the significant or unreasonable risk standard, the FDA has difficulty establishing adulteration for dietary supplement products (GAO, 2009a).

The FDA's experience with ephedra demonstrates how regulatory authority influences the agency's response to reports of safety problems. Based on numerous reports of possible adverse effects associated with the use of these products, the FDA began a process of investigating the safety of ephedra, collating and evaluating the evidence while ephedra remained on the market. Although the agency first convened an advisory committee in 1995 to investigate the adverse effects, the published final regulations banning the use of ephedra did not occur until 2004 (Yetley, 2007). Agency officials and other stakeholders attribute the difficulty of banning dietary supplements on the FDA's requirement to establish adulteration under the significant or unreasonable standard, and limited data on the safety of dietary supplements compounds the problem (GAO, 2009a).

NOTES:
[a] 21 U.S.C. § 343(r)(6).
[b] 21 U.S.C. § 343(r)(6).
[c] 21 C.F.R. § 101.93(f), (g).
[d] Public Law 109-462 (2006).
[e] Public Law 109-462 (2006).

claims in practice. Studies have demonstrated the lack of consumer understanding of the differing types of claims, leading to confusion, and in some cases, a perception that claims of lower scientific evidence are more valid than claims based on SSA. To enhance communication, more research is needed on methods of improving consumer understanding, which is a necessary prerequisite to ensuring that effective science-based strategies are used to improve consumer interpretations of health claims. Consideration should also be given to appropriate implementation of these strategies in population subgroups with differing cultural and educational backgrounds.

Several consumer studies have highlighted the difficulty of determining which strategies are most effective in influencing consumer behavior. In a 2-year effort to provide information in grocery stores on healthier food choices, Levy and colleagues (1985) found that a number of other factors were more influential on consumer purchasing behavior than the interventions tested. More influential factors included the city tested, socioeconomic status, product price, in-store purchasing trends, and seasonal trends. Levy et al. (1992) also discovered that consumers do not necessarily prefer the most effective nutrition-label formats. Mazis and Raymond (1997) found that consumers have more accurate interpretations of health claims when nutritional labels are also present, but Roe et al. (1999) found that consumers are less likely to look at Nutrition Facts panels when health claims are present on the front of food packages.

CFSAN has had a consumer studies staff for many years. This group conducts studies on consumer understanding of nutrition labeling, including health claims. There is a strong need for this type of research to be continued. These studies have informed the content of the Nutrition Facts panel through consumer studies evaluating effectiveness of various label types and consumer preferences (Heimbach and Stokes, 1982; Levy et al., 1996; Lewis and Yetley, 1992).

Additional research was conducted on consumer understanding of health claims after the introduction of qualified health claims in the marketplace. Some research showed that consumers preferred simple, succinct claim language (Williams, 2005), and other studies showed that the length and wording of claims made it difficult for consumers to identify the type of claim or strength of evidence supporting claims (Hooker and Teratanavat, 2008; Kapsak et al., 2008). The International Food Information Council Foundation (IFICF) conducted a study on consumer understanding of health claims and found that consumers rate the scientific evidence and other attributes of a product containing an authorized (SSA-level) claim similar to products containing a structure/function claim or dietary guidance statement for which FDA authorization is not required (IFICF,

2005). Research also indicates that consumers have difficulty understanding the "qualifying language" that is intended to help consumers distinguish among the four levels of scientific evidence in authorized and qualified health claims.

Both the IFICF and CFSAN reached several conclusions regarding consumer understanding of qualified health claims: first, consumers had difficulty distinguishing among the differing evidentiary levels for claims, especially with language-only claims (as opposed to graphic representations; see Figure 5-2) (CFSAN, 2005; IFICF, 2005). The IFICF study also found that words such as "promising," "inconclusive," and "may" were perceived to mean different things to different consumers, which altered their perception of the health claim (IFICF, 2005). The FDA study revealed similar perception biases: "[e]ven when qualified health claims were understood as intended, qualifying statements had unexpected effects on consumers' judgments about the health benefits and overall healthfulness of the product bearing the claim. Sometimes these qualified health claims led to more positive product perceptions" (CFSAN, 2005). Alarmingly, the FDA study also found that B grades were understood by consumers to convey greater scientific certainty than authorized health claims, or those that meet the SSA standard (in the CFSAN study, an A letter grade wasn't included for SSA health claims; instead, the substance/disease relationship was stated) (CFSAN, 2005). Further studies have suggested the need to study how consumers perceive health claims, including how the information is presented, to foster better understanding (Borra, 2006; Hooker and Teratanavat, 2008; Kapsak et al., 2008; Mazis and Raymond, 1997; Williams, 2005). Since conventional foods can make a number of claims without premarket review by the FDA (such as structure/function claims and dietary guidance statements; reviewed in Chapter 2), evidence that consumers have difficulty assessing the scientific merit of claims suggests that the multitude of claims, with differing levels of scientific support, may not adequately protect public health.

Enforcement of health claims The FDA is responsible for enforcing the correct use of food label claims; however, the general enforcement capacity of the FDA has been questioned. *Prescription for Harm: The Decline in FDA Enforcement Activity* (2006), a report requested by Rep. Henry Waxman (D-CA), found that FDA enforcement declined by more than 50 percent from 2000 to 2005. According to the report, there has been a decline in the overall number of FDA enforcement actions, including fewer warning letters and seizures. Likewise, the Center for Science in the Public Interest (CSPI) began a litigation initiative in 2005 to stop deceptive labeling, fraudulent advertising, and use of dangerous food additives, saying that these actions were necessary because the FDA and

FIGURE 5-2 Comparison of Food and Drug Administration (FDA) health-related food label statements: (A) an authorized health claim for the relationship of calcium and osteoporosis—authorized health claims require strong evidence and FDA review; (B) a nutrient content claim—these require substantiating data to be kept by the company and FDA notification but do not require FDA review; (C) a structure–function claim—these require substantiating data to be kept by the company and FDA notification but do not require FDA review; and (D) a dietary guidance statement—these are categorized separately from health claims because they make statements about healthy diet in general rather than about a specific substance in the product on which the statement appears. While claim A is based on significant scientific agreement (SSA), B, C, and D do not need to reach SSA-level evidence. Studies have indicated that consumers have difficulty understanding, or are unaware of, the levels of evidences associated with each type of label claim.
SOURCE: CFSAN (2007).

the Federal Trade Commission "have done a poor job enforcing the law in these areas" (CSPI, 2005). For example, CSPI threatened the Quaker Oats Company with a lawsuit over food labeling and advertising that "exaggerated the health benefits of eating oatmeal" (CSPI, 2007). In exchange for no longer describing oatmeal as a "unique whole grain food" that "actively finds the excess cholesterol" and a graph that overstates the cholesterol-lowering ability of oatmeal, CSPI dropped the lawsuit against the Quaker Oats Company.

However, recent enforcement actions by the FDA may indicate heightened enforcement of food labeling, including health claims. Cheerios has a cereal box that suggests "you can lower your cholesterol four percent in six weeks" (Grocery-aisle gotchas, 2009). This box refers consumers to a General Mills study that found that 3 cups of Cheerios (as opposed to 3 cups of cornflakes) with 1.5 cups of milk a day lowered cholesterol levels by 3.8 percent and low-density lipoprotein levels by 4.2 percent in six weeks (Johnston et al., 1998). Wording such as "lowers cholesterol" appears to trigger drug status, especially considering the SSA model health claim language approved by the FDA for oat bran and heart disease: "soluble fiber from foods such as oat bran in Brand Name Cereal, as part of a diet low in saturated fat and cholesterol, may reduce the risk of heart disease."[8]

On May 5, 2009, the FDA sent General Mills Inc. a warning letter over the Cheerios labeling. The FDA argued that the labeling saying "you can lower your cholesterol four percent in six weeks" violated the FDCA and applicable regulations. The letter indicated that Cheerios was being marketed as an unapproved drug and was misbranded (FDA, 2009b). A General Mills spokesperson indicated that the claim has been used for more than 2 years (Corbett Dooren, 2009). The director of CFSAN indicated that the agency is ready to send out more warning letters if it finds more violators, noting food companies have had a tendency to cross the line into the drug category by making specific health claims on packaging (Corbett Dooren, 2009).

In March 2010, the FDA notified 17 food manufacturers that the labeling for 22 of their food products violated the FDCA. The violations cited include unauthorized health claims, unauthorized nutrient content claims, and the unauthorized use of terms, such as "healthy," which have strict regulatory definitions (FDA, 2010a). In an open letter to industry, the FDA commissioner Margaret Hamburg noted that the warning letters "cover a range of concerns about how false or misleading labels can undermine the intention of Congress to provide consumers with labeling information that enables consumers to make informed and healthy food

[8] 21 C.F.R. § 101.81(e) (1997).

choices" (Hamburg, 2010). Hamburg noted that the FDA should provide as clear and consistent guidance as possible about food labeling claims and nutrition information to help consumers construct healthy diets, and that these warning letters will give the food industry further clarification about what is expected of them as they review their current labeling.

FEDERAL AGENCIES AND DATA COLLECTION

Numerous agencies within HHS are involved in efforts to collect information on medical or lifestyle interventions, observational data on health, and biomedical discovery efforts. Some of the agencies involved in data collection include the NIH, the FDA, the Centers for Disease Control and Prevention (CDC), the Centers for Medicare & Medicaid Services (CMS), and the Agency for Healthcare Research and Quality (AHRQ). In addition to the federal government, many other stakeholders are involved in the collection of data, including academia, industry, and nonprofit organizations. Government agencies and other stakeholders collect diverse information depending on the focus and needs of the organization. For example, as the steward of medical and behavioral research for the nation, the NIH conducts and supports basic and translational research. The CDC collects data to monitor health, detect and investigate health problems, to enhance prevention, and to develop and advocate sound public health policies, among its other responsibilities. In fulfilling its mission to improve the quality, safety, efficiency, and effectiveness of health care for all Americans, AHRQ conducts research that helps people make more informed decisions and improve the quality of health care services. As the nation's largest and most representative claims database, CMS collects administrative claims data, including information on diagnoses, treatment, as well as outcomes. The FDA collects information to support regulatory decision making, including the evaluation of safety and efficacy of products and ensuring that products are honestly, accurately and represented to the public.

Biomarkers are an important focus of research and data collection. Both the federal government and other stakeholders have recognized the potential role of biomarkers in the development of medical interventions, selection of populations for therapy, assessment of safety and efficacy of interventions, in clinical decision making, and in surveillance activities. Efforts to collect biomarker data are currently underway, and involve collaborations among government, industry, academic, and philanthropic stakeholders. The following section outlines ongoing efforts to collect information about biomarkers and discusses HHS' important role in facilitating and coordinating these efforts.

Collective Efforts to Collect and Share Biomarker Data

Many of the efforts to collect and share biomarker data result from precompetitive collaborations. As described by the IOM *Cancer Biomarkers* report (2007a), the challenge and expense of biomarker discovery and development may make it impossible for a company or organization to undertake the work single-handedly. Because biomarkers have the potential to facilitate research activities for multiple stakeholders, the sharing of precompetitive data and cooperation could be important to accelerating discovery and development. By pooling skills, technologies, and other resources, precompetitive collaborations may be able to leverage the strengths of different partners, leading to greater efficiency and effectiveness (IOM, 2007a). Government, industry, academia, and nonprofit organizations may potentially play roles in the sharing of precompetitive data to advance biomarker research.

The Biomarkers Consortium

For example, the Biomarkers Consortium is a public–private biomedical research partnership administered by the Foundation for the National Institutes of Health (FNIH).[9] The Consortium "endeavors to develop, validate, and qualify biological markers (biomarkers) to speed the development of medicines and therapies for detection, prevention, diagnosis, and treatment of disease and improve patient care" (FNIH, 2007). The Consortium is focused on identifying "high-impact biomarker opportunities" that address significant unmet medical needs, promise immediate practical impact on outcomes such as development of treatments and patient care, and can be accomplished within practical limits on time frames and cost. It has four disease/therapeutic focus areas: cancer, immunity and inflammation, metabolic disorders, and neuroscience. The founding partners of The Biomarkers Consortium include the NIH, the FDA, the Pharmaceutical Research and Manufacturers of America, and FNIH. Other partners include CMS and the Biotechnology Industry Organization.

[9] The FNIH was established by Congress in 1996 to support the NIH's mission of improving health through scientific discovery. According to the FNIH website, "The foundation identifies and develops opportunities for innovative public–private partnerships involving industry, academia and the philanthropic community. A nonprofit corporation, the foundation raises private-sector funds for a broad portfolio of unique programs that complement and enhance NIH priorities and activities" (FNIH, 2007). FNIH receives between $70 million and $100 million in revenues per year from benefactors such as pharmaceutical companies and the Bill & Melinda Gates Foundation.

The Biomarkers Consortium Executive Committee, composed of the founding partners and other stakeholders, has created four disease-specific Steering Committees to identify, prioritize, and refine biomarker-related project concepts and projects within their focus areas. Each Steering Committee is led by two cochairs appointed by the Executive Committee, and has broad membership, including relevant experts from key Consortium partners (NIH, FDA, CMS, industry, academia, and the advocacy community). Once the steering committee has a high-priority concept selected and refined (generally 3–4 pages in length), it will assemble a volunteer "project team" of 8–20 people with relevant expertise, including representatives from NIH, FDA, and industry. This team then develops a detailed project design and protocol (generally 30–100 pages), including a time line with expected milestones, under the direction of a staff scientific program manager. That document is then reviewed and approved by the steering committee and executive committee prior to launching a formal solicitation for funding. Most projects are launched within 4–6 months, with four to five funders supporting the study.[10]

One of the Consortium's projects includes an evaluation of the performance of adiponectin as a biomarker predictive of glycemic efficacy. Using a statistical analysis of combined data from multiple phase II clinical trials performed by four pharmaceutical companies, the project assessed the relationship between adiponectin and glucose lowering in response to PPAR (peroxisome proliferator-activated receptors) agonists. The analysis suggested that in type 2 diabetes mellitus patients, adiponectin level is a robust predictor of glycemic response to PPAR agonists, but not to non-PPAR drugs (Wagner et al., 2009). In addition, this project established important precedents for biomarker data-sharing principles among stakeholders in the Biomarkers Consortium and demonstrated the benefits of cross-company collaboration.

Under the auspices of the FNIH Biomarkers Consortium, several pharmaceutical and biotechnology industries collaborated with the National Cancer Institute (NCI), the FDA, and academic investigators to further the use of biomarkers in breast cancer treatment. The I-SPY2 (Investigation of Serial Studies to Predict Your therapeutic response with imaging and molecular analysis) trial aims to simultaneously and serially test several targeted treatments and biomarker tests to more rapidly assess which biomarkers best predict a therapeutic response (Barker et al., 2009).

[10] Personal communications with D. Wholley and S. Pearson-White, The Biomarkers Consortium, September 10, 2009.

The Critical Path Institute

The Critical Path Institute (C-Path) provides an important venue for the collection of biomarker data. C-Path, an independent, publicly-funded institute, brings together scientists from the FDA, academia, and industry to accomplish goals outlined in the Critical Path Initiative.[11] According to C-Path, its work can be viewed as a series of projects funded by grants and performed by collaborations and consortia; however, much of the organization's work falls under the goal of creating methods to enable personalized medicine that improves public health, including the creation of tools, such as biomarkers, and methods qualified by the FDA for use in medical product development (C-Path, 2010a).

As mentioned in Chapter 2, C-Path's Predictive Safety Testing Consortium (PSTC) is a public–private partnership that brings together pharmaceutical companies to share and validate each other's safety testing methods (C-Path, 2008) Through the PSTC, consortium members are sharing new preclinical biomarker tests for examination and cross-validation by other consortium members. As a result of the work of the PSTC, seven biomarkers of drug-induced nephrotoxicity in rats were validated and qualified by the FDA and the EMEA (European Medicines Agency) (FDA, 2008b).

C-Path has also initiated the Coalition Against Major Diseases (CAMD). CAMD's focus is to develop new tools and methods leading to better treatments for neurodegenerative diseases, including Alzheimer's disease and Parkinson's disease. The coalition will define clinical data standards and establish a pooled database of control arms of pharmaceutical clinical trials in order to develop quantitative disease progression models. Additionally, the coalition will attempt to incorporate imaging, biochemical, and molecular biomarkers that have the greatest potential to identify patient populations that are most likely to benefit from new therapies (C-Path, 2010b).

CEO Roundtable on Cancer's Life Science Consortium

The CEO Roundtable on Cancer's Life Science Consortium has been working to establish a new precompetitive environment to facilitate the development and use of biomarkers in cancer drug development (Curt, 2009). In recognition of the lack of standardization and qualification of biomarkers (Curt, 2009; IOM, 2007a), the Life Science Consortium envisions a new precompetitive environment that enables drug companies to

[11] As mentioned in Chapter 1, the Critical Path Initiative is the FDA's national strategy to drive innovation in the scientific processes through which FDA-regulated products are developed, evaluated, manufactured, and used (FDA, 2010b).

present their biomarker programs for cancer drug development, under confidentiality, to the NCI (Curt, 2009). This precompetitive safe harbor allows the NCI to gain a unique perspective, unobservable to its individual industry partners, to identify areas of overlap and redundancy as well as gaps. By selecting the most promising partners for further biomarker development and then sharing the validated markers with the academic and industry communities at large, the NCI provides a neutral platform that can enable cancer drug development across companies and academia because the risks are shared and collaboration replaces competition. This new approach has already come to fruition. The NCI identified a promising assay for measuring the activity of poly (ADP-ribose) polymerase inhibitors and worked to further develop and validate the assay, which has now been used in a phase 0 human trial (Kinders et al., 2008; Kummar et al., 2009; Yang et al., 2009).

Oncology Biomarker Qualification Initiative

The Oncology Biomarker Qualification Initiative (OBQI) is an agreement among the FDA, the NCI, and CMS to collaborate on improving the development of cancer therapies and the outcomes for cancer patients through biomarker development and evaluation (NCI, 2006). According to the memorandum of understanding for the OBQI collaboration, extensive cross-sector and multi-disciplinary efforts are needed to understand and develop the clinical utility of a new generation of biomarker technologies. The three agencies agreed to collaborate through working groups and steering committees to develop strategic plans, set priorities, and leverage resources and expertise from multiple sources, including the private sector (FDA, 2009g). Goals of the OBQI include the development of biomarker technologies and validation protocols to improve the detection, diagnosis, treatment, and prevention of cancer; development of guidance for the use of biomarkers to facilitate cancer drug development; and the ability to make informed decisions about reimbursement of new or existing treatment regimens based on biomarker-guided knowledge (Barker, 2006). Cancer imaging, molecular assays and targeted therapies, biomarker-driven clinical trial designs, and data mining are initial priorities of the OBQI.

Role of HHS in Biomarker Data Collection

These ongoing efforts to collect and share information about biomarkers will provide necessary information for the effective application of the committee's proposed evaluation framework. Recognizing the value of a well-coordinated, comprehensive effort to collect and share biomarker

information in advancing public health, the committee sought to improve ongoing biomarker data collection efforts. HHS can play an important role in ensuring that these efforts are optimally organized and coordinated. If biomarker discovery and development are uneven, HHS may be able to focus attention on underdeveloped areas, such as biomarkers for food and nutrition.

Unlike biomarker discovery efforts for drug and diagnostic development, relatively little research regarding biomarkers, or clinical outcomes, has been conducted for nutritional uses. One of the main reasons for this difference is that the regulatory setting for nutrition, such as foods and supplements, is quite different than for drugs and even devices. For example, drugs are generally considered unsafe and ineffective until clinical trials proving otherwise are conducted. With most foods, they are "generally recognized as safe" and can be introduced on the market without review of the safety evidence by FDA scientists. When a food manufacturer wishes to make an authorized or qualified nutritional claim about a food, the resulting health claim is not exclusive to the manufacturer, but broadly applied to the food substance across a range of food products. Other manufacturers may use an authorized health claim on their products, which decreases the incentives to collect biomarker or clinical outcome data on the food substance in their food products.

In addition, HHS coordination may ensure that biomarker data collection efforts are effective and that they leverage the stakeholders' strengths and capacities. For example, NIH has played a critical role in advancing biomarker discovery, development, and qualification (see Box 5-4). In addition, HHS can facilitate precompetitive collaborations that may encourage multiple stakeholders to share data, including industry. Such precompetitive collaborations are already underway, and include the Biomarkers Consortium, C-Path, and the OBQI. By coordinating biomarker efforts, HHS can ensure that biomarkers are effectively utilized across all contexts of use, including drugs, biologics, devices, and foods.

TRACKING THE EFFECTS OF BIOMARKER USE AT THE FDA

Within the regulatory environment, ensuring high-quality data collection is paramount to evidence-based regulation. Although the FDA has to make decisions in the presence of uncertainty, it is critical that regulatory decision making incorporates sound scientific information (Yetley, 2007). Reliance on scientific data for regulatory decision making provides legitimacy for agency actions and strengthens public trust in the FDA: "Establishing the FDA as a public health agency requires a culture that encourages scientific exchange and respects alternative viewpoints along the path of decision making. It also requires that the agency define and

BOX 5-4
Role of NIH in Biomarker Data Collection

The NIH has played an instrumental role in the development and qualification of biomarkers for all purposes. NIH has initiated a number of efforts aimed at improving collaboration between stakeholders and increasing the amount of publicly available information on promising biomarkers. These efforts include workshops on the state of the science for various biomarkers, the Biomarkers Consortium, and the Oncology Biomarker Qualification Initiative. NIH-led workshops on use of biomarkers for purposes with regulatory impact have been held via the Office of Dietary Supplements, NHLBI, and others. In 1999, the NIH and the FDA held a workshop on "Biomarkers and Surrogate Endpoints: Advancing Clinical Research and Applications" (Abstracts of the NIH-FDA conference, 1998). Topics ranged from definitions to needs and applications in disease areas from cardiovascular to psychiatric conditions. While this report does not describe the contributions of NIH and its separate institutes and offices in detail, these cannot be underestimated. The expertise, leadership, and resources of the NIH enable much rigorous science, interagency and inter-sector collaboration, and the public availability of biomarker data that would otherwise not occur. The NIH may also help play a role in prioritizing the development of biomarkers in underdeveloped areas, such as food and nutrition.

protect integrity in its basic processes" (Hamburg and Sharfstein, 2009). According to a 1998 report from the U.S. House of Representatives Committee on Science, a necessary step toward evidence-based decision making is ensuring access to sound scientific data. The report recommends that sufficient resources are committed to science that informs policy decisions so that research, whenever possible, precedes policy decisions (Committee on Science, 1998). However, there are concerns that the FDA's science capacity is at risk, threatening the agency's ability to meet current and emerging regulatory responsibilities (Subcommittee on Science and Technology, 2007).

The Subcommittee on Science and Technology concluded that: "science at the FDA is in a precarious position: the Agency suffers from serious scientific deficiencies and is not positioned to meet current or emerging regulatory responsibilities" (Subcommittee on Science and Technology, 2007). According to the subcommittee, three areas requiring improvement include strengthening mission-supportive scientific research programs, excellent staff with appropriate scientific expertise, and an information infrastructure and processing capability to ensure the FDA has access to the best data and information necessary to support regulatory science.

The subcommittee found significant deficiencies in the ability of FDA

regulatory programs to assess and use information. Although the FDA is dependent on accurate and timely information to deliver its regulatory mission, the information crisis is putting their mission at risk. The subcommittee found that there is evidence of important, but slow, progress to improve information sciences and technology at the FDA over the past few years, yet significant gaps remain. In particular, the subcommittee concluded that the FDA cannot fulfill its surveillance mission because of inadequate staff and IT resources to implement cutting-edge approaches to modeling, risk assessment, and data analysis (Subcommittee on Science and Technology, 2007). The FDA is in the process of implementing a number of initiatives to improve its capacity to collect and interpret surveillance data. The following section describes these efforts and other efforts undertaken that may be important resources to the FDA as it collects outcome data on FDA-regulated products.

Efforts to Collect Information on Outcomes

The committee recommends that the FDA ensures that appropriate data infrastructure and surveillance systems are in place to gain sufficient understanding of the effects of biomarker use. There are a number of ongoing efforts to collect information on outcomes related to FDA-regulated products. These include the Sentinel Initiative, MedWatchPlus, and the Observational Medical Outcomes Partnership. In addition, the International Serious Adverse Events Consortium, the Cardiac Safety Research Consortium, and ClinicalTrials.gov may provide important information on outcomes. Information on outcomes will need to be linked to biomarkers, so that the FDA can gain sufficient understanding of the use of biomarkers in regulatory decision making.

Sentinel Initiative

As mentioned previously, the Sentinel Initiative aims to develop and implement a proactive system to track reports of adverse events linked to the use of the FDA's regulated products (FDA, 2010c). It is hoped that the Sentinel Initiative will be a national electronic system that will transform FDA's ability to track the safety of drugs, biologics, medical devices—and ultimately all FDA-regulated products once they reach the market. The Sentinel Initiative will be developed and implemented in stages. Currently, the FDA is working on the mini-Sentinel program, or developing a Sentinel prototype. Two aspects of this prototype include developing a coordinating center for a distributed system and evaluating emerging methods in safety science (Platt, 2010). Mini-Sentinel will include drugs, biologics, and devices; data sources include administrative

claims databases, outpatient and impatient electronic medical records, and registries.

MedWatchPlus

The FDA is currently developing MedWatchPlus, an electronic system for receiving, processing, storing, and analyzing adverse event reports and other safety-related information for all FDA-regulated products. This system will combine the FDA's various safety reporting processes and systems and will provide a single point of entry for reporters. Additionally, the FDA and the NIH are collaborating to develop a "rational questionnaire" to ensure that submitting adverse events and problem reports are easier, more complete, and more consistent (FDA, 2009h). FAERS (FDA Adverse Event Reporting System) will be FDA's new repository with enhanced analytic methods to enable staff to efficiently analyze thousands of safety reports and identify potential safety problems.

Observational Medical Outcomes Partnership

The Observational Medical Outcomes Partnership (OMOP) is a public–private partnership designed to help improve the monitoring of drugs for safety. The OMOP is funded and managed through the FNIH and draws on the expertise of the FDA, other federal agencies, the pharmaceutical industry, and non-profit organizations. The partnership is conducting a 2-year initiative to research methods that are feasible and useful to analyze existing healthcare databases to identify and evaluate safety and benefit issues of drugs already on the market (FNIH, 2010). In particular, the partnership is evaluating whether multi-source observational data can improve the ability to assess drug safety and benefits (Ryan, 2010).

International SAE Consortium

The International Serious Adverse Events Consortium (iSAEC) is a nonprofit organization comprised of pharmaceutical companies, the Wellcome Trust, and academic institutions that receives scientific and strategic input from the FDA and international regulatory bodies. This consortium attempts to identify DNA variants that may be useful in predicting the risk of drug-related serious adverse events (iSAEC, 2010). The iSAEC phase 1 objectives include creation of a publicly available knowledge base of cross drug safety pharmacogenomics markers for predicting key serious adverse events and supporting the execution of the Critical Path Initiative (Holden, 2010).

Cardiac Safety Research Consortium

The Cardiac Safety Research Consortium (CSRC) is a public–private partnership of the Critical Path Initiative that focuses on cardiac safety and new medicine product development. Duke University's Clinical Research Institute manages the CSRC, which involves industry, academics, and regulators. The CSRC has developed a model for precompetitive data sharing in which electronic ECG submissions to the FDA are made available for research by the consortium, with an initial focus on QT interval issues. Additional areas of focus include using the ECG library to qualify new ECG biomarkers for cardiac risk and developing additional research and regulatory evaluation tools to facilitate clinical decision making and future medical product development (CSRC, 2010).

Clinicaltrials.gov

ClinicalTrials.gov was created in 1997 after passage of the Food and Drug Administration Modernization Act (FDAMA). As a result of FDAMA and FDAAA, Congress has required that the FDA implement registration prior to recruitment of all clinical trials that fall under the regulatory authority of the FDA and, within 1 year of completion, the reporting of results in a database. These databases have been developed by the National Library of Medicine (NLM) at the NIH in collaboration with the FDA. Drug, device, and biologic trials are included in the legislative mandates; nutritional and supplement studies are not mandated for registration or reporting, nor are observational studies covered. Nonetheless, the databases accept and encourage registration of observational studies, and about 15 percent of the registry consists of observational studies. An unknown proportion of studies are nutritional, behavioral, or health services; again, these are voluntary. Overseas trials are registered if the sponsor intends to register the drug in the United States or has U.S. study sites.

More than 80,000 trials are registered and about 500 results are available. Clinical trial results for trials initiated after September 2007 are to be provided even if the trial findings remain unpublished, as occurs in about 30 percent of trials; such reporting remains uneven. NIH and other publicly funded trials are also required to be registered and reported, but there is some confusion about the requirements. Noncompliance carries substantial penalties, so registration levels are high, with a possible exception in the device area. Individuals or researchers can use search terms that permit rapid and effective identification and aggregation. The database also links to PubMed and publications. This database has been used to scan all potential informative trials and is used in meta-analyses, which could inform qualification and validation/interpretation.

This database is built on a robust platform with public access and many links to related sources. Although the database was created and is maintained with the input of the FDA, the capability and staff necessary to administer the database reside at the NLM. Legislation is required to make changes in this reporting system, and a directive is needed regarding the implementation within the legislation. Congress has been aggressive in seeking transparency of results and their use. A major caveat is that the reporting is by investigators and sponsors, and the NLM has responsibilities for archiving, not validating, reports. Therefore, no interpretation of individual study findings is provided at the site, although the FDAAA requires an examination of whether this information can be included in an unbiased way.

Reaction to ClinicalTrials.gov has been mixed. Companies feared that disclosure of clinical trial results would put them at a competitive disadvantage and impact the viability of the pharmaceutical development enterprise (Drazen and Wood, 2005). For a period of time, some companies failed to include meaningful data in their registry entries. However, efforts on the part of the medical research community resulted in improved data submissions (Drazen and Wood, 2005, 2006). Editors of medical journals supported the database, eventually requiring authors to have registered clinical trials in the database or be barred from publication in many medical journals (Drazen and Wood, 2005). It is not yet clear how beneficial the database will be for patients and the public, due to challenges of implementation and the short time that the database has been available (Hirsch, 2008; Zarin et al., 2007).

REFERENCES

Abstracts of the NIH-FDA conference "Biomarkers and Surrogate Endpoints: Advancing Clinical Research and Applications." 1998. *Disease Markers* 14(4):187-334.

Barker, A. 2006. Oncology Biomarker Qualification Initiative: NCI-FDA-CMS collaboration to speed development of cancer therapies. Presentation to the National Cancer Advisory Board June 14, 2006. Bethesda, MD.

Barker, A., C. C. Siggman, J. K. Kelloff, N. M. Hylton, D. A. Berry, and L. J. Esserman. 2009. I-SPY 2: An Adaptive Breast Cancer Trial Design in the Setting of Neoadjuvant Chemotherapy. *Clinical Pharmacology and Therapeutics* 86(1):97–100.

Borra, S. 2006. Consumer perspectives on food labels. *American Journal of Clinical Nutrition* 83(5):1235S–1235S.

C-Path (Critical Path Institute). 2008. *Predictive Safety Testing Consortium (PSTC)*. http://www.c-path.org/pstc.cfm (accessed September 28, 2009).

C-Path. 2010a. *Sharing knowledge that leads to innovation.* http://www.c-path.org/programs.cfm (accessed March 13, 2010).

C-Path. 2010b. *Coalition Against Major Diseases.* http://www.c-path.org/CAMD.cfm (accessed March 13, 2010).

CFSAN (Center for Food Safety and Applied Nutrition). 2005. *Questions and answers: Qualified health claims in food labeling*. http://www.cfsan.fda.gov/~dms/qhc-qa.html (accessed March 17, 2009).

CFSAN. 2007. *Experimental study of health claims on food packages: Preliminary topline frequency report*. http://www.cfsan.fda.gov/~comm/crnutri4.html (accessed March 26, 2009).

CFSAN. 2008. *A food labeling guide*. http://www.cfsan.fda.gov/~dms/2lg-8.html#health (accessed March 26, 2009).

CFSAN. 2009. *Guidance for industry: Evidence-based review system for the scientific evaluation of health claims*. http://www.cfsan.fda.gov/~dms/hclmgui6.html (accessed February 25, 2009).

Coca-Cola. 2009. *2008 Annual report on form 10-K*. http://www.thecoca-colacompany.com/investors/form_10K_2008.html (accessed September 30, 2009).

Committee on Science. 1998. *Unlocking our future: Toward a new national science policy.* Washington, DC: Committee on Science, U.S. House of Representatives.

Corbett Dooren, J. 2009. Cheerios' health claims break rules, FDA says. *The Wall Street Journal*, May 13.

CSPI (Center for Science in the Public Interest). 2005. *Food watchdog group announces litigation initiative*. http://www.cspinet.org/new/200505031.html (accessed March 26, 2009).

CSPI. 2007. *Quaker agrees to tone down exaggerated health claims on oatmeal*. http://cspinet.org/new/200704171.html (accessed March 26, 2009).

CSRC (Cardiac Safety Research Consortium). 2010. *Cardiac safety: Background.* http://www.cardiac-safety.org/background (accessed March 14, 2010).

Curt, G. 2009. Step change in safe harbors: Public–private partnerships. *The Oncologist* 14(4):308–310.

Dhruva, S. S., L. A. Bero, and R. F. Redberg. 2009. Strength of study evidence examined by the FDA in premarket approval of cardiovascular devices. *Journal of the American Medical Association* 302(24):2679–2685.

Drazen, J. M., and A. J. J. Wood. 2005. Trial registration report card. *New England Journal of Medicine* 353(26):2809–2811.

Drazen, J. M., and A. J. J. Wood. 2006. Clinical trials report card. *New England Journal of Medicine* 354(13):1428–1429.

Drummond, R. 2004. Trial registration: A great idea switches from ignored to irresistible. *Journal of the American Medical Association* 292(11):1359–1362.

FDA (Food and Drug Administration). 2007. *Dietary supplement current good manufacturing practices (CGMPs) and interim final rule (IFL) facts*. http://www.fda.gov/Food/DietarySupplements/GuidanceComplianceRegulatoryInformation/RegulationsLaws/ucm110858.htm (accessed March 2, 2010).

FDA. 2008a. *FDA's mission statement*. http://www.fda.gov/opacom/morechoices/mission.html (accessed December 10, 2008).

FDA. 2008b. *FDA, European Medicines Agency to consider additional test results when assessing new drug safety*. http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/2008/ucm116911.htm (accessed November 23, 2009).

FDA. 2009a. *FDAAA implementation—Highlights two years after enactment*. http://www.fda.gov/RegulatoryInformation/Legislation/FederalFoodDrugandCosmeticActFDCAct/SignificantAmendmentstotheFDCAct/FoodandDrugAdministrationAmendmentsActof2007/ucm184271.htm (accessed November 20, 2009).

FDA. 2009b. *FDAAA implementation—Highlights two years after enactment*. http://www.fda.gov/RegulatoryInformation/Legislation/FederalFoodDrugandCosmeticActFDCAct/SignificantAmendmentstotheFDCAct/FoodandDrugAdministrationAmendmentsActof2007/ucm184271.htm (accessed November 17, 2009).

FDA. 2009c. *Food and Drug Administration Amendments Act (FDAAA) of 2007*. http://www.fda.gov/RegulatoryInformation/Legislation/FederalFoodDrugandCosmeticActFDCAct/SignificantAmendmentstotheFDCAct/FoodandDrugAdministrationAmendmentsActof2007/default.htm (accessed November 17, 2009).

FDA. 2009d. *Overview of dietary supplements*. http://www.fda.gov/Food/DietarySupplements/ConsumerInformation/ucm110417.htm#what (accessed November 6, 2009).

FDA. 2009e. *Regulatory information: Legislation*. http://www.fda.gov/RegulatoryInformation/Legislation/default.htm (accessed November 5, 2009).

FDA. 2009f. *Significant amendments to the FD&C Act*. http://www.fda.gov/RegulatoryInformation/Legislation/FederalFoodDrugandCosmeticActFDCAct/SignificantAmendmentstotheFDCAct/default.htm (accessed November 17, 2009).

FDA. 2009g. *Memorandum of understanding between the Food and Drug Administration and the National Cancer Institute and the Center for Medicare and Medicaid Services for the FDA/NCI/CMS Oncology Biomarker Qualification Initiative*. http://www.fda.gov/AboutFDA/PartnershipsCollaborations/MemorandaofUnderstandingMOUs/DomesticMOUs/ucm115681.htm (accessed March 14, 2010).

FDA. 2009h. *MedWatchPlus / FAERS.* http://www.fda.gov/ScienceResearch/SpecialTopics/CriticalPathInitiative/SpotlightonCPIProjects/ucm083295.htm (accessed March 14, 2010).

FDA. 2010a. *FDA calls on food companies to correct labeling violations; FDA commissioner issues an open letter to the industry*. http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm202814.htm (accessed March 13, 2010).

FDA. 2010b. *FDA's Critical Path Initiative: Transforming the way FDA-regulated products are developed, evaluated, manufactured and used*. http://www.fda.gov/ScienceResearch/SpecialTopics/CriticalPathInitiative/ucm076689.htm (accessed March 13, 2010).

FDA. 2010c. *FDA's Sentinel Initiative*. http://www.fda.gov/Safety/FDAsSentinelInitiative/default.htm (accessed March 14, 2010).

FNIH (Foundation for the National Institutes of Health). 2007. *The Biomarkers Consortium.* http://www.biomarkersconsortium.org/ (accessed March 13, 2010).

FNIH. 2009. *Foundation for the National Institutes of Health*. http://www.fnih.org/ (accessed November 17, 2009).

FNIH. 2010. *Observational Medical Outcomes Partnership*. http://omop.fnih.org/node/22 (accessed March 15, 2010).

Fontanarosa, P. B., D. Rennie, and C. D. DeAngelis. 2004. Postmarketing surveillance—Lack of vigilance, lack of trust. *Journal of the American Medical Association* 292(21):2647–2650.

GAO (Government Accountability Office). 2006. *Drug safety: Improvement needed in FDA's postmarket decision-making and oversight process.* Washington, DC: GAO.

GAO. 2009a. *Dietary supplements: FDA should take further actions to improve oversight and consumer understanding*. Washington, DC: GAO.

GAO. 2009b. *Food and Drug Administration: FDA faces challenges meeting its growing medical product responsibilities and should develop complete estimates of its resource needs*. Report No. GAO-09-581. Washington, DC: GAO.

GAO. 2009c. *Medical devices: Shortcomings in FDA's premarket review, postmarket surveillance, and inspections of device manufacturing establishments.* http://www.gao.gov/new.items/d09370t.pdf (accessed November 22, 2009).

GAO. 2009d. *New drug approval: FDA needs to enhance its oversight of drugs approved on the basis of surrogate endpoints*. Washington, DC: GAO.

Grocery-aisle gotchas. 2009. *Consumer Reports on Health* 21(2):8–9.

Hamburg, M. A. 2010. *Open letter to industry from Dr. Hamburg*. http://www.fda.gov/Food/LabelingNutrition/ucm202733.htm (accessed March 13, 2010).

Hamburg, M. A., and J. M. Sharfstein. 2009. The FDA as a public health agency. *The New England Journal of Medicine* 360(24):2493–2495.

Heimbach, J. T., and R. C. Stokes. 1982. Nutrition labeling and public health: Survey of American Institute of Nutrition members, food industry, and consumers. *American Journal of Clinical Nutrition* 36(4):700–708.

Hirsch, L. 2008. Trial registration and results disclosure: Impact of US legislation on sponsors, investigators, and medical journal editors. *Current Medical Research and Opinion* 24(6):1683–1689.

Holden, A. International SAE Consortium, LTD. 2010. An international industrial biomedical consortium researching the genetic basis of drug related serious adverse events. Presentation at Second Annual Sentinel Initiative Workshop. Washington, DC: The Brookings Institution.

Hooker, N. H., and R. Teratanavat. 2008. Dissecting qualified health claims: Evidence from experimental studies. *Critical Reviews in Food Science and Nutrition* 48(2):160–176.

IFICF (International Food Information Council Foundation). 2005. *Qualified health claims consumer research project executive summary*. http://www.ific.org/research/qualhealthclaimsres.cfm (accessed March 17, 2009).

IFT (Institute of Food Technologists). 2005. Functional foods: Opportunities and challenges. Chicago, IL: Institute of Food Technologists.

IOM (Institute of Medicine). 1998. *Ensuring safe food: From production to consumption*. Washington, DC: National Academy Press.

IOM. 2005. *Safe medical devices for children*. Washington, DC: The National Academies Press.

IOM. 2007a. *Cancer biomarkers: The promises and challenges of improving detection and treatment*. Washington, DC: The National Academies Press.

IOM. 2007b. *The future of drug safety: Promoting the health of the public*. Washington, DC: The National Academies Press.

IOM. 2009. *Beyond the HIPAA privacy rule: Enhancing privacy, improving health through research*. Washington, DC: The National Academies Press.

iSAEC (International Serious Adverse Events Consortium). 2010. International SAE Consortium. http://www.saeconsortium.org/ (accessed March 14, 2010).

Johnston, L., H. Reiss Reynolds, D. B. Hunninghake, K. Schultz, and B. Westereng. 1998. Cholesterol-lowering benefits of a whole grain oat ready-to-eat cereal. *Nutrition in Clinical Care* 1(1):6–12.

Kapsak, W. R., D. Schmidt, N. M. Childs, J. Meunier, and C. White. 2008. Consumer perceptions of graded, graphic and text label presentations for qualified health claims. *Critical Reviews in Food Science and Nutrition* 48(3):248–256.

Kinders, R. J., M. Hollingshead, S. Khin, L. Rubinstein, J. E. Tomaszewski, J. H. Doroshow, R. E. Parchment, and National Cancer Institute Phase 0 Clinical Trials Team. 2008. Preclinical modeling of a phase 0 clinical trial: Qualification of a pharmacodynamic assay of poly(ADP-ribose)polymerase in tumor biopsies of mouse xenografts. *Clinical Cancer Research* 14(21):6877–6885.

Kramer, D. B., E. Mallis, B. D. Zuckerman, B. A. Zimmerman, and W. H. Maisel. 2009. Premarket clinical evaluation of novel cardiovascular devices: Quality analysis of premarket clinical studies submitted to the Food and Drug Administration 2000-2007. *American Journal of Therapeutics* 17(1)2–7.

Kummar, S., R. Kinders, M. E. Gutierrez, L. Rubinstein, R. E. Parchment, L. R. Philips, J. Ji, A. Monks, J. A. Low, A. Chen, A. J. Murgo, J. Collins, S. M. Steinberg, H. Eliopoulos, V. L. Giranda, G. Gordon, L. Helman, R. Wiltrout, J. E. Tomaszewski, and J. H. Doroshow. 2009. Phase 0 clinical trial of the poly(ADP-ribose)polymerase inhibitor ABT-888 in patients with advanced malignancies. *Journal of Clinical Oncology* 27(16):2705–2711.

Levy, A. S., O. Mathews, M. Stephenson, J. E. Tenney, and R. E. Schucker. 1985. The impact of a nutrition information program on food purchases. *Journal of Public Policy & Marketing* 4(1):1–13.

Levy, A. S., S. B. Fein, and R. E. Schucker. 1992. More effective nutrition label formats are not necessarily preferred. *Journal of the American Dietetic Association* 92(10):1230–1234.

Levy, A. S., S. B. Fein, and R. E. Schucker. 1996. Performance characteristics of seven nutrition label formats. *Journal of Public Policy & Marketing* 15(1):1–15.

Lewis, C. J., and E. A. Yetley. 1992. Focus group sessions on formats on nutrition labels. *Journal of the American Dietetic Association* 92(1):62–66.

Mazis, M. B., and M. A. Raymond. 1997. Consumer perceptions of health claims in advertisements and on food labels. *Journal of Consumer Affairs* 31(1):10–26.

Meier, B. 2009. F.D.A. to seek new standards on human test data. *The New York Times*, December 30, 2009.

NCI (National Cancer Institute). 2006. *New federal health initiative to improve cancer therapy. Patients will benefit from rapid development and delivery of new cancer treatments*. http://www.cancer.gov/newscenter/pressreleases/OBQI (accessed November 17, 2009).

Office of Budget. 2009. *Fiscal Year 2010 Budget in Brief*. http://www.hhs.gov/asrt/ob/docbudget/2010budgetinbriefa.html (accessed September 30, 2009).

Office of Inspector General. 2006. *FDA's monitoring of postmarketing study commitments.* Washington, DC: Department of Health and Human Services.

Platt, R. 2010. FDA's Mini-Sentinel Program. Presentation at the Second Annual Sentinel Initiative Workshop. Washington, DC: Brookings Institution.

Platt, R. M. Wilson, A. Chan, J. S. Benner, J. Marchibroda, and M. McClellan. 2009. The new Sentinel Network—Improving the evidence of medical-product safety. *New England Journal of Medicine* 361(645-647).

Prescription for harm: The decline in FDA enforcement activity. 2006. Washington, DC: U.S. House of Representatives Committee on Government Reform—Minority Staff Special Investigations Division.

Psaty, B. M., C. D. Furberg, W. A. Ray, and N. S. Weiss. 2004. Potential for conflict of interest in the evaluation of suspected adverse drug reactions: Use of cerivastatin and risk of rhabdomyolysis. *Journal of the American Medical Association* 292(21):2622–2631.

Roe, B., A. S. Levy, and B. M. Derby. 1999. The impact of health claims on consumer search and product evaluation outcomes: Results from FDA experimental data. *Journal of Public Policy & Marketing* 18(1):89–105.

Ryan, P. 2010. Observational Medical Outcomes Partnership: Overview and lessons learned. Presentation at the Sentinel Initiative Public Workshop, Washington, DC: The Brookings Institution.

Schneeman, B. 2007. FDA's review of scientific evidence for health claims. *Journal of Nutrition* 137(2):493–494.

Subcommittee on Science and Technology. 2007. *FDA science and mission at risk*. FDA Science Board: Rockville, MD.

Taylor, C. L., and V. L. Wilkening. 2008. How the nutrition food label was developed, part 2: The purpose and promise of nutrition claims. *Journal of the American Dietetic Association* 108(4):618–623.

Wagner, J. A., E. C. Wright, M. M. Ennis, M. Prince, J. Kochan, D.J.R. Nunez, B. Schneider, M.-D. Wang, Y. Chen, S. Ghosh, B.J. Musser, and M.T. Vassileva. 2009. Utility of adinopectin as a biomarker predictive of glycemic efficacy is demonstrated by collaborative pooling of data from clinical trials conducted by multiple sponsors. *Clinical Pharmacology & Therapeutics* 86(6):619-625.

Williams, P. 2005. Consumer understanding and use of health claims for foods. *Nutrition Reviews* 63(7):256–264.

Wood, A. J. 2008. Playing "kick the FDA"—Risk-free to players but hazardous to public health. *New England Journal of Medicine* 358(17):1774–1775.

Yang, S. X., S. Kummar, S. M. Steinberg, A. J. Murgo, M. Gutierrez, L. Rubinstein, D. Nguyen, G. Kaur, A. P. Chen, V. L. Giranda, J. E. Tomaszewski, J. H. Doroshow, and The National Cancer Institute Phase 0 Working Group. 2009. Immunohistochemical detection of poly(ADP-ribose)polymerase inhibition by ABT-888 in patients with refractory solid tumors and lymphomas. *Cancer Biology and Therapy* 8(21)2004–2009.

Yetley, E. A. 2007. Science in the regulatory setting: A challenging but incompatible mix? *Novartis Foundation Symposium* 282:59–68; discussion 69–76, 212–218.

Zarin, D. A., N. C. Ide, T. Tse, W. R. Harlan, J. C. West, and D. A. Lindberg. 2007. Issues in the registration of clinical trials. *Journal of the American Medical Association* 297(19):2112–2120.

Acronyms

AAPS	American Association of Pharmaceutical Scientists
ACC	American College of Cardiology
ACE inhibitor	angiotensin-converting enzyme inhibitor
ACR	American College of Radiology
ACS	acute coronary syndrome
ACT-UP	AIDS Coalition to Unleash Power
AERS	Adverse Event Reporting System
AHA	American Heart Association
AIDS	acquired immune deficiency syndrome
BQRT	biomarker qualification review team
C-Path	Critical Path Institute
CAD	coronary artery disease
CAST	Cardiac Arrhythmia Suppression Trial
CBER	Center for Biologics Evaluation and Research
CD4 cells	$CD4^+$ T-lymphocytes
CDC	Centers for Disease Control and Prevention
CDER	Center for Drug Evaluation and Research
CDRH	Center for Devices and Radiological Health
CETP	cholesteryl ester transfer protein
C.F.R.	Code of Federal Regulations
CFSAN	Center for Food Safety and Applied Nutrition
CHD	coronary heart disease

CHF	congestive heart failure
CIN	cervical intraepithelial neoplasia
CLIA	*Clinical Laboratory Improvement Amendments*
CMOD	International Partnership for Critical Markers of Disease
CMS	Centers for Medicare & Medicaid Services
CPI	Critical Path Initiative
CRP	C-reactive protein
CRT	cardiac resynchronization therapy
CSCR	Cardiac Safety Research Consortium
CSPI	Center for Science in the Public Interest
CT	computed tomography
cTn	cardiac troponin
CVD	cardiovascular disease
DHA	docosahexaenoic acid
DIA	Drug Information Association
DRV	daily recommended value
DSHEA	Dietary Supplement Health and Education Act
DTC	direct-to-consumer
EMEA	European Medicines Agency
EPA	eicosapentaenoic acid
FAERS	FDA Adverse Event Reporting System
FDA	Food and Drug Administration
FDAAA	Food and Drug Administration Amendments Act
FDAMA	FDA Modernization Act
FDCA	Food, Drug, and Cosmetic Act
FDG-PET	[^{18}F]-2-fluoro-2-deoxy-D-glucose positron emission tomography
FNIH	Foundation for the National Institutes of Health
FOP	front of packaging
FR	*Federal Register*
FY	fiscal year
GAO	Government Accountability Office
GIST	gastrointestinal stromal tumor
GTV	gross tumor volume
HDL	high-density lipoprotein
HDL-C	high-density lipoprotein cholesterol
HHS	Department of Health and Human Services

HIV	human immunodeficiency virus
HIV-1 RNA	HIV-1 (strain of HIV) ribonucleic acid
HMG CoA	3-hydroxy-3-methylglutaryl-coenzyme A
HPV	human papillomavirus
HRT	hormone replacement therapy
Hs-CRP	high-sensitivity C-reactive protein
ICH	International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use
IFICF	International Food Information Council Foundation
IFT	Institute of Food Technologists
IOM	Institute of Medicine
IPRG	Interdisciplinary Pharmacogenomics Review Group
iSAEC	International Serious Adverse Events Consortium
ISPY-2	Investigation of Serial Studies to Predict Your Therapeutic Response with Imaging and Molecular Analysis
IT	infomation technology
LDL	low-density lipoprotein
LDL-C	low-density lipoprotein cholesterol
LDL-P	low-density lipoprotein particle number
LPS	lipopolysaccharide
LVH	left ventricular hypertrophy
MI	myocardial infarction
MRI	magnetic resonance imaging
MTHFR	5,10 methylenetetrahydrofolate reductase
NACB	National Academy of Clinical Biochemistry
NCCTG	North Central Cancer Treatment Group
NCEP	National Cholesterol Education Program
NCFST	National Center for Food Science and Technology
NCI	National Cancer Institute
NDA	New Drug Application
NHLBI	National Heart, Lung, and Blood Institute
NIH	National Institutes of Health
NLEA	Nutrition, Labeling, and Education Act
NLM	National Library of Medicine
NSABP	National Surgical Adjuvant Breast and Bowel Project
NSCLC	non-small-cell lung cancer

OBQI	Oncology Biomarker Quality Iniative
OMOP	Observational Medical Outcomes Partnership
OSE	Office of Surveillance and Epidemiology (FDA)
PDUFA	Prescription Drug User Fee Act
PET	positron emission tomography
PhRMA	Pharmaceutical Research and Manufacturers of America
PPAR	peroxisome proliferator–activated receptor
PSA	prostate-specific antigen
PSTC	Predictive Safety Testing Consortium
RACC	reference amount customarily consumed
RCT	randomized controlled trial or reverse cholesterol transport
RDI	Reference Daily Intake
REMS	risk evaluation and mitigation strategies
RiskMAPS	risk minimization action plans
RNA	ribonucleic acid
SSA	significant scientific agreement
TMUGS	tumor marker utility grading system
US	ultrasound
USDA	U.S. Department of Agriculture
VLDL	very low-density lipoprotein
VXDS	voluntary exploratory data submission
WHO	World Health Organization

Glossary

Ablation—the removal of a body part or the destruction of its function, as by a surgical procedure, morbid process, or noxious substance

Accelerated approval—regulatory mechanism by which new drugs meant to treat serious, life-threatening diseases or diseases for which there are no alternative treatments can be approved for marketing by the Food and Drug Administration using earlier clinical trial results than would be required for regular approvals; post-market surveillance and studies generally required

ACE inhibitor—see Angiotensin-converting enzyme (ACE) inhibitor

Adenomatous colon polyps—growths in the epithelial layers of the colon; can be flat, pedunculated, or sessile; result from multiple genetic mutations arising from environmental or inherited causes; can become cancerous

Adhesion molecules—molecules on cell surfaces that enable cells to stick to each other or other components of the extracellular matrix

Adjusted association—a measure of association between individual patients' true endpoints and surrogate endpoints after controlling for treatment assignment; a statistical method for surrogate endpoint evaluation

Age-related macular degeneration—a disease occurring when the cells making up a central area of the retina, called the macula, break down or move away from their normal positions; causes blurriness and sometimes loss of the center of field of vision

Analytical validation—"assessing [an] assay and its measurement performance characteristics, determining the range of conditions under which the assay will give reproducible and accurate data" (Wagner, 2002)

Angiotensin-converting enzyme (ACE) inhibitor—drug used to treat blood pressure; prevents formation of a protein that causes constriction of blood vessels, thus lowering blood pressure

Angiotension receptor blocker—type of medication used to treat high blood pressure. Unlike ACE inhibitors, which prevent the formation of angiotensin II, angiotensin receptor blockers, while allowing the protein to form, prevent it from functioning. Thus, blood pressure is lowered by preventing constriction of the blood vessels.

Apolipoprotein—a protein component of lipoprotein complexes

Arrhythmia—loss of rhythm, denoting especially an irregularity of the heartbeat

Assay—a biochemical or other measurement developed to quantitate a biomarker

Atherogenic dyslipidemia—abnormal lipid levels (including abnormal cholesterol levels) having the capacity to initiate, increase, or accelerate the process of atherogenesis having the capacity to initiate, increase, or accelerate the process of atherogenesis

Atherosclerosis/arteriosclerosis—condition characterized by irregularly distributed lipid deposits in the intima of large and medium-sized arteries; such deposits are associated with fibrosis and calcification, and are nearly always present to some degree in middle aged and older individuals

Authorized health claim—voluntary statement that characterizes the relationship between a substance and its ability to reduce the risk of disease or a health-related condition (Schneeman, 2007) that meets the significant scientific agreement (SSA) standard

Autocrine signaling—hormonal signalling in which a cell produces an agent that then binds to receptors within the same cell; related to stimulation of T-cell growth and growth of some breast cancers

Beta-carotene (β-carotene)—pigment-producing molecule in the skin of several fruits and vegetables; after ingestion, some β-carotene in bloodstream converts to two molecules of retinol (preformed vitamin A)

Bias—the systematic but unintentional erroneous association of some characteristic with a group in a way that distorts a comparison with another group (IOM, 2007)

Biological plausibility—data elucidating how the biological pathways leading from exposure to effect are useful

Biological products (biologics)—a category of products regulated by the Food and Drug Administration, including vaccines, blood and blood

components, allergenic compounds, somatic cells, gene therapy, tissues, and recombinant therapeutic proteins

Biomarker—"a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a[n]. . . intervention" (Biomarkers Definitions Working Group, 2001). Example: cholesterol level. The committee defines "objectively" to mean "reliably and accurately"

Bone mineral density (BMD)—may correlate with bone strength and a bone's ability to bear weight; may be possible to predict fracture risk using BMD as a measure

Calcium channel blocker—drug used treat heart failure caused by high blood pressure; effects the movement of calcium in the cells of the heart and blood vessels to relax blood vessels and increase the supply of blood and oxygen to the heart

Calibration—the use of measurement standards and standard measurement protocols to ensure the precision and reproducibility of an instrument or other measurement method

Cardiotoxic—having a deleterious effect on the action of the heart, due to poisoning of the cardiac muscle or of its conducting system

Cardiovascular disease—a term encompassing diseases that affect the heart and blood vessels

CD4 cell (CD4$^+$ T-cells)—specialized cells that play a role in measuring immune response in individuals with HIV

Choi criteria—a measure used to assess tumor progression in gastrointestinal stromal tumor (GIST); incorporate tumor size and tumor density into a metric of tumor progression; demonstrated to more accurately predict overall survival in GIST than reduction in tumor size

Cholesterol—abundant steroid metabolite produced by animals and found in cell membranes and circulating in blood; excess cholesterol can lead to fatty deposits in blood vessels, a risk factor for cardiovascular disease

Chromium picolinate—biologically active chromium salt that is used as a dietary supplement

Chronic disease—a culmination of a series of pathogenic processes in response to internal or external stimuli over time that results in a clinical diagnosis/ailment and health outcomes

Clinical endpoint—a characteristic or variable that reflects how a patient feels, functions, or survives (Biomarkers Definitions Working Group, 2001)

Clinical trial—a formal study carried out according to a prospectively defined protocol that is intended to discover or verify the safety and effectiveness of procedures or interventions in humans (IOM, 2007)

Clinical utility—see Utilization

Computed tomography (CT)—a special radiographic technique that uses a computer to assimilate multiple X-ray images into a two-dimensional, cross-sectional image, which also can be reconstructed into a three-dimensional image; can reveal many soft-tissue structures not shown by conventional radiography (IOM, 2007)

Congestive heart failure (CHF)—condition in which the heart is unable to maintain adequate circulation of blood in the tissues of the body or to pump out the venous blood returned to it by the venous circulation

Coronary artery disease (CAD)—see Coronary heart disease

Coronary heart disease (CHD)— refers to damage to the heart caused by atherosclerotic constriction of arteries supplying the heart; also known as coronary artery disease

C-reactive protein (CRP)—an acute-phase, non-specific, systemic marker of inflammation; in normal individuals, CRP is a trace plasma protein, but the serum concentration of CRP can increase upward of 1,000-fold upon exposure to a strong acute stimulus, such as sepsis or acute myocardial infarction

Cytostasis—the slowing of movement and accumulation of blood cells in the capillaries, as in a region of inflammation

Cytotoxic therapy—any agent or process that kills cells (e.g., chemotherapy and radiotherapy)

Diagnosis—a conclusion as to the presence of a disease

Diagnostic test—the investigative tools and techniques used in biological studies to identify or determine the presence of a disease or other condition. Any laboratory-based test that can be used in drug discovery and development as well as in patient care and clinical decision making (IOM, 2007)

Diastolic blood pressure—blood pressure as measured during the resting phase of the heart's rhythm

Dietary guidance statement—a statement describing general dietary patterns, practices and recommendations that promote health; these make reference to categories of foods and not specific substances, and they do not describe relationships between a substance (specific food or food component) and a disease or health-related condition; these can be made without Food and Drug Administration review or authorization before use

Disease—damage to an organ, part, structure, or system of the body such that it does not function properly (e.g., CHD), or a state of health leading to such dysfunctioning (e.g., hypertension)

Disease risk stratification—placement of an individual into a risk category based on the likelihood that a disease will develop or recur

Diuretic—substance promoting loss of bodily fluids through increased production and elimination of urine

Drug—materials intended for use in the diagnosis, cure, mitigation, treatment, or prevention of disease; materials (other than food) intended to affect the structure or any function of the body of humans or other animals

Early compound screening—the process of identifying compounds with the most promise for becoming efficacious and safe therapeutics

Efficacy—ability to produce a desired effect

Elliptoid model—method for estimating the volume of a cancer tumor using three different, preferably orthogonal, measurements of the tumor diameter

Epidemiologic studies—studies of the health of various human populations

Epitope—discrete site to which an antibody binds

Etiology—the science and study of the causes of disease and their mode of operation

Ex vivo—experimentation or measurements done in or on tissue in an artificial environment outside the organism

Familial hypercholesterolemia—metabolic disorder caused by defective or absent receptors for LDLs on cell surfaces; marked by an increase in blood plasma LDLs and by an accumulation of LDLs in the body resulting in xanthomas, atherosclerosis, and an increased risk of heart attack and coronary heart disease; inherited as an autosomal dominant trait

Fit-for-purpose—being guided by the principle that an evaluation process is tailored to the degree of certainty required for the use proposed

Folic acid—vitamin of the B complex that is required for normal production of red blood cells; used especially in the treatment of nutritional anemias

Food—articles used for food or drink for humans or other animals, chewing gum, and articles used for components of any such article; inclusive of foods consumed as part of meals and snacks, dietary supplements, and components contained in them (nutrients, other bioactive substances)

Friedewald formula—provides an estimate of LDL cholesterol for most fasting specimens, though its accuracy is lower at higher triglyceride concentrations.

Genomics—the study of all of the nucleotide sequences, including structural genes, regulatory sequences, and noncoding DNA segments, in the chromosomes of an organism or tissue sample. One example of the application of genomics in oncology is the use of microarray or other techniques to uncover the genetic "fingerprint" of a tissue sample. This

genetic fingerprint is the pattern that stems from the variable expression of different genes in normal and cancer tissues (IOM, 2007)

Glycation—the uncontrolled, non-enzymatic reaction of sugars with proteins; important in the damage done to diabetics when their sugar levels rise above normal, and in damage done to critical proteins of long-lived nerve cells in aging

Glycosylated hemoglobin—hemoglobin to which glucose is bound; tested to monitor the long-term control of diabetes mellitus

Growth factor—a substance (e.g., vitamin B12 or an interleukin) that promotes cellular growth

Health claim—a claim that describes the relationship between a substance (food or food component) and a disease or health-related condition; limited to claims about disease risk reduction and cannot be claims about the cure, mitigation, treatment, or prevention of disease

Heart disease—*see* Cardiovascular disease

Hemostasis regulator—biological chemical involved in the process of stopping blood flow, as from a broken blood vessel

High-density lipoprotein cholesterol (HDL-C)—a lipoprotein of blood plasma that is composed of a high proportion of protein with little triglyceride and cholesterol and that is associated with decreased probability of developing atherosclerosis

Hill criteria—criteria used to establish cause in the case of non-infectious or chronic disease by evaluating strength, consistency, specificity, temporality, biological gradient, plausibility, coherence, experiment, analogy; used in environmental health, toxicology, pharmacology, epidemiology, and medicine

Hyperlipidemia—the presence of abnormally high amount of lipids in the circulating blood

Hypertension—abnormally high arterial blood pressure that is usually indicated by an adult systolic blood pressure of 140 mm Hg or greater or a diastolic blood pressure of 90 mm Hg or greater; can result in thickening and inelasticity of arterial walls and damage to the heart; a risk factor for various pathological conditions or events (e.g., heart attack, heart failure, stroke, end-stage renal disease, or retinal hemorrhage)

Immune response—response of an organism to a foreign substance

Immunoassay—test measuring the immune response of an organism to an antigen

In vitro—outside the living body and in an artificial environment

In vivo—in the living body of a plant or animal

Incidence—occurrence of a disease or condition

Inflammatory biomarker—biological chemicals indicating tissue damage or irritation; C-reactive protein is an inflammatory biomarker

Inflammatory response (inflammation)—a local response to cellular injury that is marked by capillary dilatation, leukocytic infiltration, redness, heat, pain, swelling, or loss of function and that serves as a mechanism initiating the elimination of elimination of foreign substances and for healing damaged tissue

Insulin resistance—an organism's inability to respond to and use the insulin it produces; this condition is related to the type 2 diabetes incidence

Intermediate endpoint—a biologic event or marker that is a precursor to a given health outcome (e.g., atherosclerosis for cardiovascular disease endpoints or blurred vision for macular degeneration)

Intervention—any drug, device, biologic, behavioral modification, nutritional modification, lifestyle modification, or other treatment intended to improve health

Intima-media thickness (IMT)—the thickness of the inner layers of an artery

Ischemic stroke—stroke caused by thrombosis or embolism; caused by an inadequate flow of blood to heart tissue due to a constriction or blockage to blood vessels supplying it

Lipoprotein—compounds containing lipid and protein; almost all lipids in plasma are lipoproteins

Low-density lipoprotein cholesterol (LDL-C)—a lipoprotein of blood plasma that is composed of a moderate proportion of protein with little triglyceride and a high proportion of cholesterol and that is associated with increased probability of developing atherosclerosis

Luminal structures—relating to the lumen of a blood vessel

Macromolecules—large molecules, often polymeric or with colloidal properties; examples include many proteins, nucleic acids, and polysaccharides

Magnetic resonance imaging (MRI)—method by which images are created by recording signals generated from the excitation (the gain and loss of energy) of such elements as the hydrogen of water in tissue when placed in a powerful magnetic field and pulsed with radiofrequencies (IOM, 2007)

Mass spectrometry—a method for separating ionized molecular particles according to mass by applying a combination of electrical and magnetic fields to deflect ions passing in a beam through the instrument (IOM, 2007)

Medical device—any instrument, apparatus, appliance, material, or other article intended to be used to affect the structure or any function of a human or animal body

Mendelian randomization—the random assignment of genetic material from parents to offspring; a tool used in epidemiology to help deter-

mine whether a health outcome is caused by genetic or environmental factors or to elucidate gene–gene or gene–environment interactions

Metabolomics—the systematic study of the unique chemical fingerprints that specific cellular processes leave behind, that is, small-molecule metabolites (IOM, 2007)

Microarray—a high-throughput biological assay in which different probes are deposited on a chip surface (glass or silicon) in a miniature arrangement; DNA microarrays most commonly used (IOM, 2007)

Myocardial infarction—an acute episode of heart disease marked by the death or damage of heart muscle due to insufficient blood supply to the heart muscle, usually as a result of a coronary thrombosis or a coronary occlusion and that is characterized especially by chest pain

Myxoid degeneration—a degenerative process in which the connective tissues are replaced by a gelatinous or mucoid substance

Neural tube effects—a group of birth defects that involve the central nervous system; result from failure of the neural tube to properly form

Normal sinus rhythm—normal heart rhythm

Null hypothesis—the hypothesis that an intervention has no effect (i.e., that there is no true difference in outcomes between a treatment group and a control group); typically, if statistical tests indicate that the P-value is at or above the specified a-level (e.g., 0.01 or 0.05), then any observed treatment effect is not statistically significant, and the null hypothesis cannot be rejected

Nutrient content claim—statements about the level of a nutrient or dietary substance in the product, using terms such as free, high, and low, or they compare the level of a nutrient in a food to that of another food, using terms such as *more*, *reduced*, and *lite*

Opportunity cost—for any decision, the loss of the benefits of the next best alternative decision

Oxidation—chemical reaction between a substance and oxygen. Fire and rust are examples of oxidative processes.

P-value—a measure of the probability that a subsequent measurement's magnitude would be equal or greater to the measured magnitude if the null hypothesis is true—in other words, if there is no true difference between the control and experimental groups

Paracrine signal—referring to the release of locally acting substances from endocrine cells

Paraneoplastic—caused by or resulting from the presence of cancer in the body, but not the physical presence of cancerous tissue in the part or organ affected

Pathogenesis—the mode of origin or development of any disease or morbid process

Pathophysiology—processes leading to the incidence or progression of disease or other health-related condition; alteration in function as distinguished from structural defects

Patient selection—in clinical trials, patient selection (inclusion/exclusion) by disease subset or probability of response/adverse events

Peripheral vascular disease—a type of cardiovascular disease caused by atherosclerosis of the arteries to the limbs, reducing the blood supply and therefore depriving the limb muscles of oxygen

Pharmacodynamic assay—a test used to determine a drug's activity; can be used to select dose quantities and schedule

Pharmacologic response—the effect of a drug on an organism in relation to the concentration of the drug

Phase I trial—clinical trial in a small number of patients in which the toxicity and dosing of an intervention are assessed (IOM, 2007)

Phase II trial—clinical trial in which the safety and preliminary efficacy of an intervention are assessed in patients (IOM, 2007)

Phase III trial—large-scale clinical trial in which the safety and efficacy of an intervention are assessed in a large number of patients. The Food and Drug Administration generally requires new drugs to be tested in phase III trials before they can be put on the market (IOM, 2007)

Phospholipid—any of numerous lipids in which phosphoric acid as well as a fatty acid is esterified to glycerol and which are found in all living cells and in the bilayers of cell membranes

Phytosterol—cholesterol-like compounds found in vegetable oils, nuts, and legumes; may reduce serum cholesterol

Plaque—a well-demarcated yellow area or swelling on the surface of the artery; produced by intimal lipid deposit

Plasma—the fluid portion of the circulating blood

Pleiotropic effects—having multiple phenotypic expressions; for example, the non-lipid effects of statins, including the anti-inflammatory and antithrombotic properties that contribute to an improvement in vascular function

Polymorphism—occurrence in the same population of two or more genotypes of such proportion that the most rare cannot be maintained by recurrent mutation alone; heritable variations in low-density lipoproteins; variant lipoproteins exhibit different antigenic and chemical properties compared with normal lipoproteins

Positive harm—something that is intended to do good is not only ineffective, but causes definite harm as an unintended side effect

Positive predictive value—the probability that an individual with a positive test has, or will develop, a particular disease, or characteristic, that the test is designed to detect; a measure of the ratio of true positives to (false + true positives) (IOM, 2007)

Positron emission tomography (PET)—a highly sensitive technique that uses radioactive probes to image in vivo tumors, receptors, enzymes, DNA replication, gene expression, antibodies, hormones, drugs, and other compounds and processes (IOM, 2007)

Postmarket studies—may be mandated by the Food and Drug Administration for already approved drugs or devices to review potential risks

Precision—a measure of random error; inversely related to random error; confidence intervals are computed to demonstrate the precision of relative risk estimates

Predictive value—the ability to predict the change in the outcome of a disease given a particular intervention using a specified patient measurement

Prentice criteria—stringent requirements to be met before a biomarker can definitively substitute for a clinical endpoint for a given use; briefly, the criteria state that a biomarker must perfectly correlate with the clinical outcome it is meant to replace and capture the entire effect of the intervention used to bring about the effect on the clinical outcome

Prevalence (disease)—the number of existing cases of a disease in a given population at a specific time

Prevention—the use of medical and public health tools to prevent disease, injury, or other events injurious to health

Prognosis—an assessment of the probable course of a disease given the risk factors present in an individual; this assessment may affect treatment decisions

Prognostic value—the ability to predict disease outcome or course using a specified patient measurement

Protease—biological chemical that reacts with proteins, degrading them chemically and making them non-functional

Proteomics—the study of the structure, function, and interactions of the proteins produced by the genes of a particular cell, tissue, or organism. The application of proteomics in oncology may involve mass spectroscopy, two-dimensional polyacrylamide gel electrophoresis, protein chips, and other techniques to uncover the protein "fingerprint" of a tissue sample. This protein fingerprint is the pattern that stems from the various amounts and types of all the proteins in the sample (IOM, 2007)

Qualification—evidentiary process of linking a biomarker with biological processes and clinical endpoints

Qualified health claim—voluntary statement that characterizes the relationship between a substance and its ability to reduce the risk of disease or a health-related condition (Schneeman, 2007) that does not meet the significant scientific agreement (SSA) standard

Randomized controlled trial (RCT)—a study in which the participants are assigned by chance to separate groups that compare different treatments; neither the researchers nor the participants can choose which group. Using chance to assign people to groups means that the groups will be similar and that the treatments they receive can be compared objectively. At the time of the trial, it is not known which treatment is best

Relative effect—the effect of a treatment on the distribution of true endpoints versus surrogate endpoints; component of a statistical method for surrogate endpoint evaluation

Relative risk—the ratio of the risk of disease in exposed individuals to the risk of disease in non-exposed individuals

Risk–benefit analysis—the comparison of the risk of a situation to its benefits

Risk biomarker—biomarker that indicates a component of an individual's level of risk for developing a disease or level of risk for developing complications of a disease

Risk stratification—the classification of patients into groups based on the likelihood of developing or suffering effects from a disease

Safety biomarker—a biomarker that can be used to identify patients at high risk for serious side effects, to monitor early signs of toxicity, or to predict the likelihood for severe toxicity

Sample bias—see Bias

Saturated fat— fat having no double bonds; chemically the most stable type of fat; solid at room temperature; come chiefly from animal food products; tend to raise the level of cholesterol in the blood

Screening—the use of risk factor analysis and biomarker assays to detect early-stage disease in the asymptomatic population

Sensitivity (analytical)—the lowest concentration that can be distinguished from background noise; this concentration is termed an assay's detection limit (IOM, 2007)

Sensitivity (clinical)—a measure of how often a test correctly identifies patients with a specific diagnosis. It is calculated as the number of true-positive results divided by the number of true-positive plus false-negative results (IOM, 2007).

Sepsis—the presence of various pathogenic organisms, or their toxins, in the blood or tissues

Serum—the fluid portion of the blood obtained after removal of fibrinogen, other clottting factors, and cells; a clear watery fluid, especially that moistening surface of serous membranes

Significant scientific agreement (SSA)—judgment that qualified experts would likely agree that the scientific evidence supports the substance–disease relationship that is the subject of a proposed health claim

Specificity (analytical)—how well an assay detects only a specific substance and does not detect closely related substances (IOM, 2007)

Specificity (clinical)—a measure of how often a test correctly identifies the proportion of persons without a specific diagnosis; calculated as the number of true-negative results divided by the number of true-negative plus false-positive results (IOM, 2007)

Statin—any of a group of drugs (as lovastatin and simvastatin) that inhibit the synthesis of cholesterol and promote the production of LDL-binding receptors in the liver, resulting in a decrease in the level of LDL and a modest increase in the level of HDL circulating in blood plasma

Structure–function claim—statements describing the role of a nutrient or dietary ingredient intended to affect normal structure or function in humans; may characterize the means by which a nutrient or dietary ingredient acts to maintain such structure or function; may describe general well-being from consumption of a nutrient or dietary ingredient; manufacturer is responsible for ensuring the accuracy and truthfulness of the statement; FDA does not review these claims prior to manufacturer use

Substance—a specific food (tomato) or component of food (lycopene), whether in conventional food or dietary supplement form

Supplement—a product taken by mouth that contains a dietary ingredient intended to supplement the diet; dietary ingredients may include: vitamins, minerals, herbs or other botanicals, amino acids, and substances such as enzymes, organ tissues, glandulars, and metabolites; may be found in forms such as tablets, capsules, softgels, gelcaps, liquids, or powders

Surrogate endpoint—a biomarker that is intended to substitute for a clinical endpoint. A surrogate endpoint is expected to predict clinical benefit (or harm or lack of benefit or harm) based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence (IOM, 2007)

Surrogate threshold effect—the minimum treatment effect on the surrogate endpoint necessary to predict a non-zero effect on the true endpoint; provides information relevant to the practical use of a surrogate endpoint and can be interpreted from a clinical point of view

Surveillance—population-level monitoring for early detection and treatment of advancing disease or complications

Systolic blood pressure—the highest blood pressure that occurs during a beat of the heart, just after the left ventricle has contracted

Target validation—demonstration that a potential drug target plays a key role in the disease process

Therapeutic intervention—actions taken (through administration of a drug or other means) to treat a disease or other health-related condition

Therapy monitoring—the process of determining whether a therapy is having the intended effect on a disease and whether adverse effects arise

Total cholesterol—total amount of cholesterol (both LDL and HDL) in the blood

Toxicology—the science of understanding the effects of chemicals on humans and other organisms

Triglyceride—any of a group of lipids that are esters formed from one molecule of glycerol and three molecules of one or more fatty acids, are widespread in adipose tissue, and commonly circulate in the blood in the form of lipoproteins

Troponin—protein of muscle that together with tropomyosin forms a regulatory protein complex controlling the interaction of actin and myosin and that when combined with calcium ions permits muscular contraction (e.g., of the heart)

True endpoint—the endpoint for which a surrogate endpoint is sought

Tumor response rates—in its most primitive form: tumor shrinkage; defined by a change in tumor bulk; commonly used for making decisions regarding approval of anticancer drugs in the 1970s

Tumor size—inconsistently defined biomarker often used for determining efficacy of cancer therapeutics

Type 2 diabetes—diabetes mellitus of a common form that develops especially in adults and most often in obese individuals and that is characterized by hyperglycemia resulting from impaired insulin utilization coupled with the body's inability to compensate with increased insulin production

Ultracentrifuge—a high-speed centrifuge by means of which large molecules (proteins, nucleic acids) are caused to sediment at practicable rates; used for determination of molecular weights

Utilization—contextual analysis based on the specific use proposed and the applicability of available evidence to this use. This includes a determination of whether the validation and qualification conducted provide sufficient support for the use proposed

Validation—see Analytical validation

Vasodilator—an agent that causes dilation of the blood vessels

Ventricular tachycardia—relatively rapid heart action (whether physiological or pathological) that is associated with the generation of electrical impulses within the ventricles and is characterized by an electrocardiogram having a broad QRS complex. A QRS complex is a measurable characteristic of an electrocardiogram

Viral fitness—refers to the relative replication competence of a virus (e.g., HIV) under defined circumstances; generally assessed in tissue culture systems; its relevance to the clinical situation may be difficult to fully establish

REFERENCES

Biomarkers Definitions Working Group. 2001. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology and Therapeutics* 69(3):89–95.

IOM (Institute of Medicine). 2007. *Cancer biomarkers: The promises and challenges of improving detection and treatment*. Washington, DC: The National Academies Press.

Schneeman, B. 2007. FDA's review of scientific evidence for health claims. *Journal of Nutrition* 137(2):493–494.

Wagner, J. A. 2002. Overview of biomarkers and surrogate endpoints in drug development. *Disease Markers* 18(2):41–46.

Appendix A

Table of Papers About Biomarker Qualification

TABLE A-1 Historical Review of the Biomarker–Surrogate Endpoint Literature with Special Reference to the Nomenclature, Initial Reports, Systems of Classification, and Statistical Methods Developed for Their Evaluation

Year	Author	Focus	Field/Summary and Commentary
1963	Mainland	Nomenclature: Substituted variables	Statistics and Medicine In his *Elementary Medical Statistics*, he discusses substituting variables that are easy to observe for ones that are difficult to observe.
1966	Rushing	Nomenclature: First report of surrogate used in any context	Psychology, Ethics, Social Science, Law The role of the hospital nurse as a mother surrogate. (Many publications followed in the 1960s and 1970s where surrogate was used in this context of a person's role in the fields of psychology, ethics, social science, and law.)
1973	Rho et al.	Nomenclature: First report of biomarker	Biology A search for porphyrin biomarkers in nonesuch shale and extraterrestrial samples. *Biomarker* here represents biological marker—origins of biological life.
1976	Schlenger	Nomenclature: First report of surrogate AND outcome	Epidemiology Mortality and morbidity rates as surrogates for "health."
1977	Karpetsky et al.	Nomenclature: Second report of biomarker	Oncology Serum RNase level was found to be an indicator of renal function, and was not a biomarker either for the presence or extent of the plasma cell tumor. (Forty of 46 biomarker reports from 1977 to 1985 were in oncology.)
1978	Baker	Nomenclature: Third report of biomarker	Oncology Preoperative assessment of the patient with breast cancer.
1980	Regelson	Nomenclature: First report of biomarker outside cancer in medicine	General Medicine Biomarkers in aging: A beginning for a therapeutic approach in Transactions of the Association of Life Insurance Medical Directors of America.

TABLE A-1 Continued

Year	Author	Focus	Field/Summary and Commentary
	Webb and Lin	Nomenclature: First report of biomarker in title of publication	Oncology Urinary fibronectin: Potential as a biomarker in prostatic cancer.
1982	Waalkes et al.	Biomarkers for clinical application	Oncology Feasibility study in the development of 17 biological markers for ovarian cancer.
1983	Wood	Nomenclature: First report of surrogate AND endpoint, second report of surrogate AND outcome	Rheumatology Nature of surrogate endpoints. Relationships considered at two levels: (1) ability of the attribute to act as a surrogate in detection of the underlying state (at a particular point in time); (2) potential of the surrogate to reveal changes in the underlying state as its course unfolds.
1986	Bigger	Second surrogate and endpoint, third surrogate and outcome	Cardiology Electrophysiological testing to select patients with ventricular arrhythmias for drug trials and to determine anti-arrhythmic drug efficacy. (By the end of the decade, the use of biomarkers as surrogates in cardiology had a number of high-profile failures.)
	Buccheri et al.	First report of biomarker as measure of tumor burden and predict outcome	Oncology Clinical value of a multiple biomarker assay (CEA, TPA, b-HCG, LDH) in patients with bronchogenic carcinoma.
1987	Kalish et al.	Third surrogate and endpoint	Oncology Surrogates as endpoints in bladder cancer trials. Data show that superficial disease endpoints do not predict surrogates for invasive disease endpoints.
	Schulof et al.	Surrogates markers first used as response to therapy	HIV Phase I/II trial of thymosin fraction 5 and thymosin alpha one in HTLV-III–seropositive subjects.

continued

TABLE A-1 Continued

Year	Author	Focus	Field/Summary and Commentary
	Rosin et al.	Intermediate endpoints	Oncology Promise of intermediate endpoints in quantitating the response of precancerous lesions to chemopreventive agents.
1989	The Cardiac Arrhythmia Suppression Trial (CAST) Investigators; Ruskin	First example of study to test surrogate; treatment of a biomarker successful, but patient outcome worse	Cardiology CAST showed that successful suppression of the ventricular arrhythmia biomarker with antiarrhythmic therapy was associated with increased rather than decreased patient mortality.
	Herson	First substantive discussion on surrogate endpoints in clinical trials	Methodology An introduction to four invited papers on surrogate endpoints in clinical trials. These were pivotal papers. Trigger was an FDA criticism of new drug applications in cardiology and oncology because they used surrogate endpoints.
	Ellenberg and Hamilton	All key issues discussed using examples from oncology	Methodology Advantages and disadvantages of surrogate endpoints. Key points: Used when endpoints of interest are too difficult and/or expensive to measure; must be sufficiently well correlated with the endpoints of interest to justify substitution; initial choice often based on biologic rationale as primary endpoints are more acceptable in early drug development than later pivotal studies.
	Wittes et al.	Many key issues discussed using examples from cardiology	Methodology Key points: "True" endpoint is one with clinical importance to the patient, such as mortality or a major clinical outcome; surrogate is one biologically closer to the process of disease; surrogate is useful if easily measured and highly correlated with the true endpoint; surrogates can dramatically reduce sample size and trial duration.

TABLE A-1 Continued

Year	Author	Focus	Field/Summary and Commentary
	Prentice	First report addressing the key statistical barrier to the use of surrogates	Statistics Prentice defines a surrogate endpoint to be a "response variable for which a test of the null hypothesis of no relationship to the treatment groups under comparison is also a valid test of the corresponding null hypothesis based on the true endpoint."
	Beaudry and Spence	First report of surrogate outcome	Cardiology Atherosclerosis severity index based on noninvasive ultrasound assessment to replace angiographic measurement of atherosclerosis (costly and invasive), which in turn replaced clinical endpoints (latter most expensive). (Example of developing a surrogate to replace another surrogate.)
	Buchwald et al.	Empirical surrogate endpoint validation	Cardiology RCT to demonstrate a reduction in overall mortality by lipid modification and to validate coronary arteriographic change as a surrogate for change in coronary heart disease risk.
1990	Machado et al.	Testing validity of surrogate therapeutics	HIV Medicine Pros for surrogate endpoints: Ethical/practical reasons for hastening decision making about the efficacy of new treatments for HIV infection. Cons: Serious overestimates of clinical benefit if treatment had delayed toxicity or only transient beneficial effects; serious underestimates of clinical benefit when the treatment had no effect on the transition from healthy to the marker state.

continued

TABLE A-1 Continued

Year	Author	Focus	Field/Summary and Commentary
	Schatzkin et al.	Statistical validation strategy	Oncology The intermediate endpoint is a valid cancer surrogate if the attributable proportion is near 1.0, but not if it is near 0 (usually the attributable proportion is neither 1.0 nor 0); in this case in an established exposure-cancer relationship, the exposure effect would vanish if adjusted for the intermediate endpoint.
	Woosley	Further commentary on CAST results and implications for drug development	Cardiology High-profile study that illustrated the dangers of surrogate therapeutics. (Further failures followed in other cardiology studies. Within a few years, surrogates rarely used in cardiology and large outcome trials with patient endpoints were the norm. Other fields in medicine did not have resources to conduct large, long studies and continued to argue for the use of surrogates in drug development.)
	Lippman et al.	Schema	Oncology Proposed three classes of biomarkers: genomic, proliferation, and differentiation markers. Biomarker validation studies should follow an evolutionary process. This leads to first generation (short-term trials in high-risk patients), second generation (dose and schedule trials), and third generation trials (long-term phase III trials to validate first generation candidate biomarkers).
1992	New drug, antibiotic, and biological drug product regulations; accelerated approval—FDA. Final Rule[a]	FDA "accelerated approval" regulation	General Accelerate approval of new drugs and biological products for serious or life-threatening illnesses, with provisions for any necessary continued study of the drugs' clinical benefits after approval.

TABLE A-1 Continued

Year	Author	Focus	Field/Summary and Commentary
	Freedman et al.	Statistical validation	Statistics Statistical validation of intermediate endpoints requires exposure or intervention effect, adjusted for the intermediate endpoint, to be reduced to zero. The estimating statistic—PTE—is explained by the intermediate/surrogate endpoint and its 95% confidence limits are determined.
	Boissel et al.	Schema	Methodology Three provisos for surrogate outcome evaluation. Proviso 1, the surrogate endpoint, should occur more frequently than corresponding clinical endpoint. Proviso 2, that relationship between the surrogate and clinical endpoint, is well established through relevant epidemiological studies. Proviso 3, that the estimate of the expected clinical benefit should be derivable from the estimate of the reduction on the surrogate endpoint, which can be obtained from randomized clinical trials data.
	Freedman et al.	Schema	Methodology A new validation criterion based on an analysis of the three-way relationship of exposure (E), marker (M), and disease (D). Provides the level of evidence required for using intermediate markers as endpoints for Phase II and Phase III trials. (These criteria were conceptual and qualitative only.)
	Freedman and Schatzkin	Sample-size issues	Methodology Different sample-size requirements for questions on surrogate endpoint validity: Does the intervention affect the intermediate endpoint? Is the intermediate endpoint associated with the main outcome? Is the intervention effect on the main outcome mediated by the intermediate endpoint?

continued

TABLE A-1 Continued

Year	Author	Focus	Field/Summary and Commentary
1993	The Hypertension Optimal Treatment Study (the HOT Study)	Targeting biomarker	Cardiology Dose–response relationship between surrogate target and clinical outcome.
	Lin et al.	Application of Prentice	AIDS CD4-lymphocyte count captures part of the relationship between zidovudine and time to a first critical event, but does not fulfill the Prentice criterion.
1994	Aickin	Surrogate endpoint biomarker	Oncology *If there is gold in the labeling index hills, are we digging in the right place*? (Tool for cancer chemoprevention studies.)
1995	Temple	Schema	Methodology "Feels function or survives" definition for surrogate endpoint.
	Lee et al.	Review	Methodology Surrogate biochemical markers: Precise measurement for strategic drug and biologics development.
	Hughes et al.	Review	Statistics/HIV Evaluating surrogate markers.
	Scientific Advisory Committee on Surrogate Markers of HIV	Consensus	HIV Medicine Consensus statement. Scientific advisory committee on surrogate markers of HIV.

TABLE A-1 Continued

Year	Author	Focus	Field/Summary and Commentary
1996	Fleming and DeMets	Review	Methodology Surrogate endpoints in clinical trials. Are we being misled? Argues for use of surrogate endpoints in Phase II, but not Phase III pivotal trials. Failure of surrogate endpoints because: (1) surrogate is not in the causal pathway of the disease process; (2) of several causal pathways of the disease, the intervention affects only the pathway mediated through the surrogate; (3) surrogate is not in the pathway of the intervention's effect or is insensitive to its effect; and (4) intervention has mechanisms of action independent of disease process.
	Schatzkin et al.	Review	Methodology Surrogate endpoints in cancer research: a critique.
1997	De Gruttola et al.	Schema	Methodology Validating surrogate markers: Are we being naïve? The variety of proposed metrics for evaluating the degree to which this criterion is met are subject to misinterpretation because of the multiplicity of mechanisms by which drugs operate. Without detailed understanding of these mechanisms, metrics of "surrogacy" are not directly interpretable. Even when all of the mechanisms are understood, these metrics are associated with a high degree of uncertainty unless either treatment effects are large in moderate-sized studies or sample sizes are large in studies of moderately effective treatments.
	Lin et al.	Statistics	Statistics Estimating the proportion of a treatment effect explained by surrogate marker.

continued

TABLE A-1 Continued

Year	Author	Focus	Field/Summary and Commentary
	Mildvan et al.	Schema	Methodology An approach to the validation of markers for use in AIDS clinical trials.
	Rolan	Schema	Methodology The contribution of clinical pharmacology surrogates and models to drug development. Proposes five dimensional properties of surrogates. These are validation (statistical), innovation, proximity to clinical outcome, specificity for an intervention, and practicality.
	Topol et al.	Review	Methodology Need clinical endpoints to establish safety and efficacy.
	Daniels and Hughes	Schema	Statistical Method/HIV Meta-analysis for the evaluation of potential surrogate markers.
	Boissel et al.	Schema	Methodology Clinical evaluation: From intermediate to surrogate criteria (French).
	Colburn	Schema	Methodology Selecting and validating biologic markers for drug development.
1998	Albert et al.	Review–consensus	Methodology/HIV Statistical issues for HIV surrogate endpoints: Point/counterpoint.
	Buyse and Molenberghs	Statistics	Statistics Introduction of the relative effect (RE) and adjusted association (AA) for single-unit studies.
	FDA and NIH	Review and abstracts	Methodology Biomarkers and surrogate endpoints: Advancing clinical research and applications. (Abstracts.)

TABLE A-1 Continued

Year	Author	Focus	Field/Summary and Commentary
	Hayes	Schema	Methodology Tumor Marker Utility Grading System (TMUGS) proposed to evaluate the clinical utility of tumor markers and to establish an investigational agenda for evaluation of new tumor markers for risk assessment, screening, differential diagnosis, prognosis, monitoring clinical course, and use in clinical trials. Includes a TMUGS Worksheet that clarifies the precise characteristics of the marker in question and evaluates its clinical utility on a six-point scale (ranging from 0 to +++).
1999	Bucher et al.	Schema	Methodology How to use and article measuring the effects of an intervention on surrogate endpoints.
2000	Buyse et al.	Statistics	Statistics Validation of surrogate endpoints in meta-analysis of randomized experiments.
	Buyse et al.	Statistics	Statistics Statistical validation of surrogate endpoints.
	Colburn	Schema	Methodology Optimizing the use of biomarkers, surrogate endpoints, and clinical endpoints for efficient drug development.

continued

TABLE A-1 Continued

Year	Author	Focus	Field/Summary and Commentary
	Gail et al.	Statistics	Methodology The strengths and weakness of meta-analytic assessment of surrogate endpoints: (1) which trials? (2) how many trials? (3) difficult to obtain individual-level data to estimate within study variance. (4) between-study variation can yield much less precise estimates of treatment effects on true-endpoint than estimates based on true-endpoint itself. (5) realistic models for distribution complicated. (6) difficulty modeling joint or marginal distributions of true-endpoint and surrogate. (7) which approach frequentist, empirical Bayes, and Bayesian for hierarchical systems. (8) how to use covariates. (9) unanticipated toxicity. Conclusion: Meta-analysis of surrogate endpoints may lead to less precise estimates of treatment effect on clinical endpoint than relying on clinical endpoint itself.
	Begg and Leung	Statistics	Statistics Provide conceptual alternatives to Prentice criterion for surrogate statistical validation.
	Schatzkin	Review	Methodology Intermediate markers as surrogate endpoints in cancer research.
	Fleming	Review	Methodology Brief review of practical and statistical issues.
2001	Li et al.	Statistics	Statistics A method to assess the proportion of treatment effect explained by a surrogate endpoint—a general model and graphical setting.
	Lesko and Atkinson	Schemas	Methodology Biomarkers and surrogate endpoints in drug development and regulatory decision making.

TABLE A-1 Continued

Year	Author	Focus	Field/Summary and Commentary
	Xu and Zeger	Statistics	Statistics Evaluation of multiple surrogate endpoints.
	Biomarkers Definitions Working Group	Schema	Methodology Biomarkers and surrogate endpoints: preferred definitions and conceptual framework.
	De Gruttola et al.	Schema	Methodology Considerations and recommendations in evaluation of surrogate endpoints in clinical trials: Summary of NIH workshop.
2002	Wang and Taylor	Statistics	Statistics A measure of the proportion of treatment effect explained by a surrogate marker.
	Lathia	Review	Methodology Biomarkers and surrogate endpoints: How and when might they impact drug development?
	Molenberghs et al.	Statistics	Statistics Statistical challenges in the evaluation of surrogate endpoints in randomized trials.
	Wagner	Review	Methodology Overview of biomarkers and surrogate endpoints in drug development.
	Cowles	Statistics	Statistics Bayesian estimation of the PTE captured by a surrogate marker.
	Schatzkin and Gail	Review	Methodology Promise and peril of surrogate endpoints in cancer research: Review of the logical issues as well as the problem of measurement error.
	Frangakis and Rubin	Statistics	Statistics Principal stratification in causal inference.
	Henderson et al.	Statistics	Statistics Longitudinal modeling.

continued

TABLE A-1 Continued

Year	Author	Focus	Field/Summary and Commentary
	Lin et al.	Statistics	Statistics Latent class models for joint analysis.
	Taylor and Wang	Statistics	Statistics Surrogate markers and joint models.
	Hughes	Comment	Methodology Imprecision in the estimates require modeling.
2003	Rolan et al.	Review	Methodology Use of biomarkers from drug discovery through clinical practice. Mechanistic classification into six types of biomarkers.
	Baker and Freedman	Statistics	Statistics Method for analyzing data from a randomized trial with a missing binary outcome.
	Baker and Kramer	Review	Methodology A perfect correlate does not a surrogate make.
2004	FDA	Position paper	FDA's Critical Path Document *Innovation or Stagnation: Challenge and Opportunity on the Critical Path to New Medical Products*.
	Berger	Statistics	Statistics Does Prentice criterion validate surrogate endpoints?
	Molenberghs et al.	Statistics	Methodology Perspective of surrogate endpoints in controlled trials.
	Alonso et al.	Statistics	Methodology Role of statistics in surrogate endpoints.
	Rubin	Statistics	Methodology Direct versus indirect causal effects.

TABLE A-1 Continued

Year	Author	Focus	Field/Summary and Commentary
	Baker et al.	Review	Drug Development A general framework for describing various roles for biomarkers in cancer prevention research (early detection, surrogate endpoint, and cohort identification for primary prevention) and the phases in their evaluation.
2005	Fleming	Review	Methodology Surrogate endpoints and FDA's accelerated approval process.
	Sargent et al.	Statistics	Oncology Meta-analytic approach for surrogate validation.
	Baker, 2006a	Statistics	Methodology A simple meta-analytic approach for using a binary endpoint to predict the effect of intervention on true endpoint.
	Korn et al.	Statistics	Methodology Assessing surrogates as trial endpoints using mixed models.
2006	Weir and Walley	Statistics	Review Statistical evaluation of biomarkers as surrogate endpoints: A literature review.
	Baker, 2006b	Statistics	Review Title: Surrogate endpoints: Wishful thinking or reality?
	Finley Austin and Babiss	Review	Methodology Where and how could biomarkers be used in 2016?
	Qu and Case	Statistics	Statistics Quantifying the indirect treatment effect via surrogate markers.
	Desai et al.	Review	Cardiology Blood pressure as an example of a biomarker that functions as a surrogate.
	Hughes	Review	HIV Medicine Initial treatment of HIV Infection: Randomized trials with clinical endpoints are still needed.

continued

TABLE A-1 Continued

Year	Author	Focus	Field/Summary and Commentary
	Johnson et al.	Statistics	Oncology Prediction bands used in a meta-analysis of RCTs to determine the surrogate threshold for response-rate and time to progression endpoints as predictors of mortality in metastatic colorectal cancer and non-small-cell lung cancer.
	FDA	Regulatory initiatives	Update on Critical Path Initiative.
2007	Lassere et al., 2007b	Schema	Methodology Definitions and validation criteria for biomarkers and surrogate endpoints: Development and testing of quantitative hierarchical levels of evidence schema.
	Lassere et al., 2007a	Statistics	Review Simulation studies of surrogate endpoint validation using single trial and multitrial statistical approaches.
	Wagner et al.	Schema	Methodology Biomarker qualification, a graded, "fit-for-purpose" qualitative evidentiary process linking a biomarker with biology and clinical endpoints.

NOTES: [a] 57 *Federal Register* 239 (1992) pp. 58942–58960. AA = adjusted association; AIDS = acquired immune deficiency syndrome; b-HCG = beta-human chorionic gonadotropin; CAST = The Cardiac Arrhythmia Suppression Trial; CEA = carcinoembryonic antigen; FDA = Food and Drug Administration; HIV = human immunodeficiency virus; HOT = The Hypertension Optimal Treatment Study; HTLV-III = human T-lymphotropic virus type III; LDH = lactate dehydrogenase; NIH = National Institutes of Health; PTE = proportion of treatment effect; RCT = randomized controlled trial; RE = relative effect; TPA = tissue plasminogen activator.

SOURCE: Lassere (2008). Reprinted with permission from SAGE publications, Copyright 2009 by SAGE Publications.

TABLE A-2 Continuation of Table A-1 for 2007-2009

Year	Author	Focus	Field/Summary and Commentary
2007	Alonso and Molenberghs	Statistics	Methodology An information-based validation method for surrogate endpoints.
	Pryseley et al.	Statistics	Methodology The authors test and review a meta-analytic approach to biomarker qualification and support use of a recently proposed, more computationally efficient process in some circumstances.
	Rasnake et al.	Regulatory	Nutrition Discussion of emerging surrogate endpoints and the use of surrogate endpoints in the review of health claims at the FDA.
2008	Alonso and Molenberghs	Statistics	Methodology/Oncology Evaluation of time to cancer recurrence as a surrogate endpoint for survival, as evaluated using a meta-analytic framework.
	Altar et al.	Schema	Methodology Provides an "evidence map" for grading available evidence and a process for biomarker qualification.
	Burzykowski	Comment	Methodology A concise summary of the topic of surrogate endpoint qualification.
	Chakravarty and Sridhara	Regulatory	Oncology/Regulatory Issues Discussion of use of progression-free survival as a trial endpoint.
	Green et al.	Statistics	Methodology Use of multiple methods, both statistical and clinically relevant qualitative methods, is proposed.
	Joy and Hegele	Comment	Methodology Discussion of the failure of the torcetrapib trials and the implications for CETP inhibition as a treatment target.
	Krumholz and Lee	Comment	Methodology Recent failures of surrogate endpoints in cardiology and endocrinology.

continued

TABLE A-2 Continued

Year	Author	Focus	Field/Summary and Commentary
	Lassere	Review, schema	Methodology Systematic review of biomarker and surrogate endpoint validation criteria from 1950 to 2007; also provides criteria for ranking surrogate validity.
	Osborne	Comment	Alzheimer's/Regulatory Comment on shifts in use of surrogate endpoints for Alzheimer's disease drug development.
	Psaty and Lumley	Comment	Cardiology Further discussion of recent surrogate endpoint failures in lipid-altering drug clinical trials.
	Wagner	Schema	Methodology Comprehensive discussion of fit-for-purpose biomarker qualification for all stages of drug development.
2009	Hlatky et al.	Schema	Methodology/Cardiology Title: Criteria for evaluation of novel markers of cardiovascular risk: A scientific statement from the American Heart Association.
	Lathia et al.	Schema	Methodology Successes and failures in use of surrogate endpoints for drug development; discussion of necessary criteria for surrogate endpoint qualification and use.
	Prentice	Statistics	Methodology Title: Surrogate and mediating endpoints: Current status and future directions.
	Rigatto and Barrett	Review	Methodology Statement of definitions, advantages, and disadvantages to biomarker and surrogate endpoint use for clinical trials.

TABLE A-2 Continued

Year	Author	Focus	Field/Summary and Commentary
	Shi and Sargent	Review	Methodology Discussion of surrogate endpoints evaluation and the use of meta-analysis of multiple clinical trials for evaluation.

NOTES: CETP = cholesteryl ester transfer protein; FDA = Food and Drug Administration.

REFERENCES

Aickin, M. 1994. If there is gold in the labeling index hills, are we digging in the right place? *Journal of Cellular Biochemistry Supplement* 19:91–93.

Albert, J. M., J. P. A. Ioannidis, P. Reichelderfer, B. Conway, R. W. Coombs, L. Crane, R. Demasi, D. O. Dixon, P. Flandre, M. D. Hughes, L. A. Kalish, K. Larntz, D. Y. Lin, I. C. Marschner, A. Munoz, J. Murray, J. Neaton, C. Pettinelli, W. Rida, J. M. G. Taylor, S. L. Welles, and NIAID Workshop. 1998. Statistical issues for HIV surrogate endpoints: Point/counterpoint. *Statistics in Medicine* 17(21):2435–2462.

Alonso, A., and G. Molenberghs. 2007. Surrogate marker evaluation from an information theory perspective. *Biometrics* 63(1):180–186.

Alonso, A., and G. Molenberghs. 2008. Surrogate end points: Hopes and perils. *Expert Review of Pharmacoeconomics and Outcomes Research* 8(3):255–259.

Alonso, A., G. Molenberghs, T. Burzykowski, D. Renard, H. Geys, Z. Shkedy, F. Tibaldi, J. C. Abrahantes, and M. Buyse. 2004. Prentice's approach and the meta-analytic paradigm: A reflection on the role of statistics in the evaluation of surrogate endpoints. *Biometrics* 60(3):724–728.

Altar, C. A., D. Amakye, D. Buonos, J. Bloom, G. Clack, R. Dean, V. Devanarayan, D. Fu, S. Furlong, L. Hinman, C. Girman, C. Lathia, L. Lesko, S. Madani, J. Mayne, J. Meyer, D. Raunig, P. Sager, S. A. Williams, P. Wong, and K. Zerba. 2008. A prototypical process for creating evidentiary standards for biomarkers and diagnostics. *Clinical Pharmacology and Therapeutics* 83(2):368–371.

Baker, R. R. 1978. Preoperative assessment of patient with breast-cancer. *Surgical Clinics of North America* 58(4):681–691.

Baker, S. G. 2006a. A simple meta-analytic approach for using a binary surrogate endpoint to predict the effect of intervention on true endpoint. *Biostatistics* 7(1):58–70.

Baker, S. G. 2006b. Surrogate endpoints: Wishful thinking or reality? *Journal of the National Cancer Institute* 98(8):502–503.

Baker, S. G., and L. S. Freedman. 2003. A simple method for analyzing data from a randomized trials with a missing binary outcome. *BMC Medical Research Methodology* 3:8.

Baker, S. G., and B. S. Kramer. 2003. A perfect correlate does not a surrogate make. *BMC Medical Research Methodology* 3:16.

Baker, S. G., B. S. Kramer, and P. C. Prorok. 2004. Development tracks for cancer prevention markers. *Disease Markers* 20(2):97–102.

Beaudry, M., and J. D. Spence. 1989. Measurement of atherosclerosis—Development of an atherosclerosis severity index. *Clinical and Experimental Hypertension Part a–Theory and Practice* 11(5–6):943–956.

Begg, C. B., and D. H. Y. Leung. 2000. On the use of surrogate end points in randomized trials. *Journal of the Royal Statistical Society Series a-Statistics in Society* 163:15–24.

Berger, V. W. 2004. Does the Prentice criterion validate surrogate endpoints? *Statistics in Medicine* 23(10):1571–1578.

Bigger Jr., J. T. 1986. Long-term continuous electrocardiographic recordings and electrophysiologic testing to select patients with ventricular arrhythmias for drug trials and to determine antiarrhythmic drug efficacy. *American Journal of Cardiology* 58(5):58C–65C.

Biomarkers Definitions Working Group. 2001. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology and Therapeutics* 69(3):89–95.

Boissel, J. P., J. P. Collet, P. Moleur, and M. Haugh. 1992. Surrogate endpoints: A basis for a rational approach. *European Journal of Clinical Pharmacology* 43(3):235–244.

Boissel, J. P., L. Perret, G. Bouvenot, A. Castaigne, M. J. GerardCoue, P. Maillere, P. Mismetti, M. Vray, B. Avouac, B. Begaud, C. Caulin, Y. Charpak, P. ChaumetRiffaud, D. Costagliola, F. Degos, J. M. Destors, F. Doyon, F. Fagnani, I. Giri, J. P. Girre, J. M. Goehrs, A. Kher, D. Moccatti, A. Munoz, A. P. deCurzon, Y. Pletan, P. Roy, D. Vasmant, and A. Wajman. 1997. Intermediate and surrogate outcomes in clinical evaluation of drugs. Therapie 52(4):281–285.

Buccheri, G. F., B. Violante, A. M. Sartoris, D. Ferrigno, A. Curcio, and F. Vola. 1986. Clinical-value of a multiple biomarker assay in patients with bronchogenic-carcinoma. *Cancer* 57(12):2389–2396.

Bucher, H. C., G. H. Guyatt, D. J. Cook, A. Holbrook, and F. A. McAlister. 1999. Users' guides to the medical literature. XIX. Applying clinical trial results. A. How to use an article measuring the effect of an intervention on surrogate end points. *Journal of the American Medical Association* 282(8):771–778.

Buchwald, H., J. P. Matts, L. L. Fitch, R. L. Varco, G. S. Campbell, M. Pearce, A. Yellin, R. D. Smink, H. S. Sawin, C. T. Campos, B. J. Hansen, and J. M. Long. 1989. Program on the surgical control of the hyperlipidemias (POSCH)—Design and methodology. *Journal of Clinical Epidemiology* 42(12):1111–1127.

Burzykowski, T. 2008. Surrogate endpoints: Wishful thinking or reality? *Statistical Methods in Medical Research* 17(5):463–466.

Buyse, M., and G. Molenberghs. 1998. Criteria for the validation of surrogate endpoints in randomized experiments. *Biometrics* 54(3):1014–1029.

Buyse, M., G. Molenberghs, T. Burzykowski, D. Renard, and H. Geys. 2000a. Statistical validation of surrogate endpoints: Problems and proposals. *Drug Information Journal* 34(2):447–454.

Buyse, M., G. Molenberghs, T. Burzykowski, D. Renard, and H. Geys. 2000b. The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics* 1(1):49–67.

The Cardiac Arrhythmia Suppression Trial (CAST) Investigators. 1989. Preliminary report: Effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. The Cardiac Arrhythmia Suppression Trial (CAST) Investigators. *New England Journal of Medicine* 321(6):406–412.

Chakravarty, A., and R. Sridhara. 2008. Use of progression-free survival as a surrogate marker in oncology trials: Some regulatory issues. *Statistical Methods in Medical Research* 17(5):515–518.

Colburn, W. A. 1997. Selecting and validating biologic markers for drug development. *Journal of Clinical Pharmacology* 37(5):355–362.

Colburn, W. A. 2000. Optimizing the use of biomarkers, surrogate endpoints, and clinical endpoints for more efficient drug development. *Journal of Clinical Pharmacology* 40(12 Pt 2):1419–1427.

Cowles, M. K. 2002. Bayesian estimation of the proportion of treatment effect captured by a surrogate marker. *Statistics in Medicine* 21(6):811–834.

Daniels, M. J., and M. D. Hughes. 1997. Meta-analysis for the evaluation of potential surrogate markers. *Statistics in Medicine* 16(17):1965–1982.

De Gruttola, V., T. Fleming, D. Y. Lin, and R. Coombs. 1997. Perspective: Validating surrogate markers—Are we being naive? *Journal of Infectious Diseases* 175(2):237–246.

De Gruttola, V. G., P. Clax, D. L. DeMets, G. J. Downing, S. S. Ellenberg, L. Friedman, M. H. Gail, R. Prentice, J. Wittes, and S. L. Zeger. 2001. Considerations in the evaluation of surrogate endpoints in clinical trials: Summary of a National Institutes of Health Workshop. *Controlled Clinical Trials* 22(5):485–502.

Desai, M., N. Stockbridge, and R. Temple. 2006. Blood pressure as an example of a biomarker that functions as a surrogate. *AAPS Journal* 8(1):E146.

Ellenberg, S. S., and J. M. Hamilton. 1989. Surrogate endpoints in clinical trials: Cancer. *Statistics in Medicine* 8(4):405–413.

FDA (Food and Drug Administration). 2004. *Challenge and opportunity on the critical path to new medical products*. http://www.fda.gov/oc/initiatives/criticalpath/whitepaper.html (accessed December 10, 2008).

FDA. 2006. *Critical Path opportunities list*. http://www.fda.gov/oc/initiatives/criticalpath/reports/opp_list.pdf (accessed December 10, 2008).

FDA and NIH (National Institutes of Health). 1999. Abstracts of the NIH–FDA conference "Biomarkers and surrogate endpoints: Advancing clinical research and applications." *Disease Markers* 14(4):187–334.

Finley Austin, M. J., and L. Babiss. 2006. Commentary: Where and how could biomarkers be used in 2016? *AAPS Journal* 8(1):E185–E189.

Fleming, T. R. 2000. Surrogate end points in cardiovascular disease trials. *American Heart Journal* 139(4):S193–S196.

Fleming, T. R. 2005. Surrogate endpoints and FDA's Accelerated Approval Process: The challenges are greater than they seem. *Health Affairs* 24(1):67–78.

Fleming, T. R., and D. L. DeMets. 1996. Surrogate end points in clinical trials: Are we being misled? *Annals of Internal Medicine* 125(7):605–613.

Frangakis, C. E., and D. B. Rubin. 2002. Principal stratification in causal inference. *Biometrics* 58(1):21–29.

Freedman, L. S., and A. Schatzkin. 1992. Sample-size for studying intermediate end-points within intervention trials or observational studies. *American Journal of Epidemiology* 136(9):1148–1159.

Freedman, L. S., B. I. Graubard, and A. Schatzkin. 1992. Statistical validation of intermediate endpoints for chronic diseases. *Statistics in Medicine* 11(2):167–178.

Gail, M. H., R. Pfeiffer, H. C. Van Houwelingen, and R. J. Carroll. 2000. On meta-analytic assessment of surrogate outcomes. *Biostatistics* 1(3):231–246.

Green, E. M., G. Yothers, and D. J. Sargent. 2008. Surrogate endpoint validation: Statistical elegance versus clinical relevance. *Statistical Methods in Medical Research* 17(5):477–486.

Hayes, D. F. 1998. Determination of clinical utility of tumor markers: A tumor marker utility grading system. *Recent Results in Cancer Research* 152:71–85.

Henderson, R., P. Diggle, and A. Dobson. 2002. Identification and efficacy of longitudinal markers for survival. *Biostatistics* 3(1):33–50.

Herson, J. 1989. The use of surrogate endpoints in clinical trials (an introduction to a series of four papers). *Statistics in Medicine* 8(4):403–404.

Hlatky, M. A., P. Greenland, D. K. Arnett, C. Ballantyne, M. H. Criqui, M. S. V. Elkind, A. S. Go, F. E. Harrell Jr., B. V. Howard, V. J. Howard, P. Y. Hsue, C. M. Kramer, J. P. McConnell, S. T. Normand, C. J. O'Donnell, J. Smith, S. C., and P. W. F. Wilson. 2009. Criteria for evaluation of novel markers of cardiovascular risk. A scientific statement from the American Heart Association. *Circulation* 119:2408–2416.

Hughes, M. D. 2002. Evaluating surrogate endpoints. *Controlled Clinical Trials* 23(6): 703–707.

Hughes, M. D. 2006. Initial treatment of HIV infection: Randomized trials with clinical end points are still needed. *Journal of Infectious Diseases* 194(5):542–544.

Hughes, M. D., V. De Gruttola, and S. L. Welles. 1995. Evaluating surrogate markers. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* 10(Suppl 2):S1–S8.

Hypertension Optimal Treatment Study (the HOT Study). 1993. *Blood Pressure* 2(1):62–68.

Johnson, K. R., C. Ringland, B. J. Stokes, D. M. Anthony, N. Freemantle, A. Irs, S. R. Hill, and R. L. Ward. 2006. Response rate or time to progression as predictors of survival in trials of metastatic colorectal cancer or non-small cell lung cancer: A meta-analysis. *Lancet Oncology* 7(9):741–746.

Joy, T. R., and R. A. Hegele. 2008. The failure of torcetrapib: What have we learned? *British Journal of Pharmacology* 154(7):1379–1381.

Kalish, L. A., M. B. Garnick, and J. P. Richie. 1987. Appropriate end-points for superficial bladder-cancer clinical-trials. *Journal of Clinical Oncology* 5(12):2004–2008.

Karpetsky, T. P., R. L. Humphrey, and C. C. Levy. 1977. Influence of renal-insufficiency on levels of serum ribonuclease in patients with multiple-myeloma. *Journal of the National Cancer Institute* 58(4):875–880.

Korn, E. L., P. S. Albert, and L. M. McShane. 2005. Assessing surrogates as trial endpoints using mixed models. *Statistics in Medicine* 24(2):163–182.

Krumholz, H. M., and T. H. Lee. 2008. Redefining quality—Implications of recent clinical trials. *New England Journal of Medicine* 358(24):2537–2539.

Lassere, M. N. 2008. The biomarker-surrogacy evaluation schema: A review of the biomarker-surrogate literature and a proposal for a criterion-based, quantitative, multidimensional hierarchical levels of evidence schema for evaluating the status of biomarkers as surrogate endpoints. *Statistical Methods in Medical Research* 17(3):303–340.

Lassere, M., K. Johnson, M. Hughes, D. Altman, M. Buyse, S. Galbraith, and G. Wells. 2007a. Simulation studies of surrogate endpoint validation using single trial and multitrial statistical approaches. *Journal of Rheumatology* 34(3):616–619.

Lassere, M. N., K. R. Johnson, M. Boers, P. Tugwell, P. Brooks, L. Simon, V. Strand, P. G. Conaghan, M. Ostergaard, W. P. Maksymowych, R. Landewe, B. Bresnihan, P.-P. Tak, R. Wakefield, P. Mease, C. O. Bingham III, M. Hughes, D. Altman, M. Buyse, S. Galbraith, and G. Wells. 2007b. Definitions and validation criteria for biomarkers and surrogate endpoints: Development and testing of a quantitative hierarchical levels of evidence schema. *Journal of Rheumatology* 34(3):607–615.

Lathia, C. D. 2002. Biomarkers and surrogate endpoints: How and when might they impact drug development? *Disease Markers* 18(2):83–90.

Lathia, C. D., D. Amakye, W. Dai, C. Girman, S. Madani, J. Mayne, P. MacCarthy, P. Pertel, L. Seman, A. Stoch, P. Tarantino, C. Webster, S. Williams, and J. A. Wagner. 2009. The value, qualification, and regulatory use of surrogate end points in drug development. *Clinical Pharmacology and Therapeutics* 86(1):32–43.

Lee, J. W., J. D. Hulse, and W. A. Colburn. 1995. Surrogate biochemical markers—Precise measurement for strategic drug and biologics development. *Journal of Clinical Pharmacology* 35(5):464–470.

Lesko, L. J., and J. Atkinson, A. J. 2001. Use of biomarkers and surrogate endpoints in drug development and regulatory decision making: Criteria, validation, strategies. *Annual Review of Pharmacology and Toxicology* 41:347–366.

Li, Z. Q., M. P. Meredith, and M. S. Hoseyni. 2001. A method to assess the proportion of treatment effect explained by a surrogate endpoint. *Statistics in Medicine* 20(21):3175–3188.

Lin, D. Y., M. A. Fischl, and D. A. Schoenfeld. 1993. Evaluating the role of CD4-lymphocyte counts as surrogate end-points in human-immunodeficiency-virus clinical-trials. *Statistics in Medicine* 12(9):835–842.

Lin, D. Y., T. R. Fleming, and V. DeGruttola. 1997. Estimating the proportion of treatment effect explained by a surrogate marker. *Statistics in Medicine* 16(13):1515–1527.

Lin, H. Q., B. W. Turnbull, C. E. McCulloch, and E. H. Slate. 2002. Latent class models for joint analysis of longitudinal biomarker and event process data: Application to longitudinal prostate-specific antigen readings and prostate cancer. *Journal of the American Statistical Association* 97(457):53–65.

Lippman, S. M., J. S. Lee, R. Lotan, W. Hittelman, M. J. Wargovich, and W. K. Hong. 1990. Biomarkers as intermediate end points in chemoprevention trials. *Journal of the National Cancer Institute* 82(7):555–560.

Machado, S. G., M. H. Gail, and S. S. Ellenberg. 1990. On the use of laboratory markers as surrogates for clinical endpoints in the evaluation of treatment for HIV infection. *Journal of Acquired Immune Deficiency Syndrome* 3(11):1065–1073.

Mainland, D. 1963. *Elementary medical statistics*. Philadelphia, PA: W. B. Saunders.

Mildvan, D., A. Landay, V. De Gruttola, S. G. Machado, and J. Kagan. 1997. An approach to the validation of markers for use in AIDS clinical trials. *Clinical Infectious Diseases* 24(5):764–774.

Molenberghs, G., M. Buyse, H. Geys, D. Renard, T. Burzykowski, and A. Alonso. 2002. Statistical challenges in the evaluation of surrogate endpoints in randomized trials. *Controlled Clinical Trials* 23(6):607–625.

Molenberghs, G., T. Burzykowski, A. Alonso, and M. Buyse. 2004. A perspective on surrogate endpoints in controlled clinical trials. *Statistical Methods in Medical Research* 13(3):177–206.

Osborne, R. 2008. FDA and surrogate endpoints in Alzheimer's disease. *Nature Biotechnology* 26(8):843.

Prentice, R. L. 1989. Surrogate endpoints in clinical trials: Definition and operational criteria. *Statistics in Medicine* 8(4):431–440.

Prentice, R. L. 2009. Surrogate and mediating endpoints: Current status and future directions. *Journal of the National Cancer Institute* 101(4):216–217.

Pryseley, A., A. Tilahun, A. Alonso, and G. Molenberghs. 2007. Information-theory based surrogate marker evaluation from several randomized clinical trials with continuous true and binary surrogate endpoints. *Clinical Trials* 4(6):587–597.

Psaty, B. M., and T. Lumley. 2008. Surrogate end points and FDA approval: A tale of 2 lipid-altering drugs. *Journal of the American Medical Association* 299(12):1474–1476.

Qu, Y., and M. Case. 2006. Quantifying the indirect treatment effect via surrogate markers. *Statistics in Medicine* 25(2):223–231.

Rasnake, C. M., P. R. Trumbo, and T. M. Heinonen. 2007. Surrogate endpoints and emerging surrogate endpoints for risk reduction of cardiovascular disease. *Nutrition Reviews* 66(2):76–81.

Regelson, W. 1980. Biomarkers in aging: A beginning for a therapeutic approach. *Transactions of the Association of Life Insurance Medical Directors of America* 64:170–184.

Rho, J. H., A. J. Bauman, H. G. Boettger, and T. F. Yen. 1973. A search for porphyrin biomarkers in Nonesuch Shale and extraterrestrial samples. *Space Life Sciences* 4(1):69–77.

Rigatto, C., and B. J. Barrett. 2009. Biomarkers and surrogates in clinical studies. In *Clinical epidemiology: Practice and methods*, Methods in Molecular Biology Vol. 473, edited by B. Barrett and P. Parfrey. Secaucus, NJ: Springer. Pp. 137–154.

Rolan, P. 1997. The contribution of clinical pharmacology surrogates and models to drug development—A critical appraisal. *British Journal of Clinical Pharmacology* 44(3):219–225.

Rolan, P., A. J. Atkinson Jr., L. J. Lesko, Scientific Organizing Committee, and Conference Report Committee. 2004. Use of biomarkers from drug discovery through clinical practice. Report of the Ninth European Federation of Pharmaceutical Sciences Conference on Optimizing Drug Development. *Clinical Pharmacology & Therapeutics* 73(4):284–291.

Rosin, M. P., B. P. Dunn, and H. F. Stich. 1987. Use of intermediate end-points in quantitating the response of precancerous lesions to chemopreventive agents. *Canadian Journal of Physiology and Pharmacology* 65(3):483–487.

Rubin, D. B. 2004. Direct and indirect causal effects via potential outcomes. *Scandinavian Journal of Statistics* 31(2):161–170.

Rushing, W. A. 1966. The hospital nurse as mother surrogate and bedside psychologist. A sociological appraisal. *Mental Hygiene* 50(1):71–80.

Ruskin, J. N. 1989. The Cardiac Arrhythmia Suppression Trial (CAST). *New England Journal of Medicine* 321(6):386–388.

Sargent, D. J., H. S. Wieand, D. G. Haller, R. Gray, J. K. Benedetti, M. Buyse, R. Labianca, J. F. Seitz, C. J. O'Callaghan, G. Francini, A. Grothey, M. O'Connell, P. J. Catalano, C. D. Blanke, D. Kerr, E. Green, N. Wolmark, T. Andre, R. M. Goldberg, and A. De Gramont. 2005. Disease-free survival versus overall survival as primary end point for adjuvant colon cancer studies: Individual patient data from 20,898 patients on 18 randomized trials. *Journal of Clinical Oncology* 23(34):8664–8670.

Schatzkin, A. 2000. Intermediate markers as surrogate endpoints in cancer research. *Hematology/Oncology Clinics of North America* 14(4):887–905.

Schatzkin, A., and M. Gail. 2002. The promise and peril of surrogate endpoints in cancer research. *Nature Reviews Cancer* 2(1):19–27.

Schatzkin, A., L. S. Freedman, M. H. Schiffman, and S. M. Dawsey. 1990. Validation of intermediate end points in cancer research. *Journal of the National Cancer Institute* 82(22):1746–1752.

Schatzkin, A., L. S. Freedman, J. Dorgan, L. M. McShane, M. H. Schiffman, and S. M. Dawsey. 1996. Surrogate end points in cancer research: A critique. *Cancer Epidemiology Biomarkers & Prevention* 5(12):947–953.

Schlenger, W. E. 1976. A new framework for health. *Inquiry* 13(3):207–214.

Schulof, R. S., G. L. Simon, M. B. Sztein, D. M. Parenti, R. A. DiGioia, J. W. Courtless, J. M. Orenstein, C. M. Kessler, P. D. Kind, S. Schlesselman, H. M. Paxton, M. Robert-Guroff, P. H. Naylor, and A. L. Goldstein. 1986. Phase I/II trial of thymosin fraction 5 and thymosin alpha one in HTLV-III seropositive subjects. *Journal of Biological Response Modifiers* 5(5):429–443.

Scientific Advisory Committee on Surrogate Markers of HIV. 1995. Consensus statement. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* 10(Suppl 2): S114–S116.

Shi, Q., and D. J. Sargent. 2009. Meta-analysis for the evaluation of surrogate endpoints in cancer clinical trials. *International Journal of Clinical Oncology* 14(2):102–111.

Taylor, J. M., and Y. Wang. 2002. Surrogate markers and joint models for longitudinal and survival data. *Controlled Clinical Trials* 23(6):626–634.

Temple, R. J. 1995. A regulatory authority's opinion about surrogate endpoints. In *Clinical measurement in drug evaluation*, edited by W. S. Nimmo and G. T. Tucker. New York: John Wiley and Sons. Pp. 3–22.

Topol, E. J., R. M. Califf, F. VandeWerf, M. Simoons, J. Hampton, K. L. Lee, H. White, J. Simes, and P. W. Armstrong. 1997. Perspectives on large-scale cardiovascular clinical trials for the new millennium. *Circulation* 95(4):1072–1082.

Waalkes, T. P., J. E. Mrochek, S. R. Dinsmore, and D. C. Tormey. 1978. Serum protein-bound carbohydrates for following course of disease in patients with metastatic breast carcinoma. *Journal of the National Cancer Institute* 61(3):703–707.

Wagner, J. A. 2002. Overview of biomarkers and surrogate endpoints in drug development. *Disease Markers* 18(2):41–46.

Wagner, J. A. 2008. Strategic approach to fit-for-purpose biomarkers in drug development. *Annual Review of Pharmacology and Toxicology* 48:631–651.

Wagner, J. A., S. A. Williams, and C. J. Webster. 2007. Biomarkers and surrogate end points for fit-for-purpose development and regulatory evaluation of new drugs. *Clinical Pharmacology and Therapeutics* 81(1):104–107.

Wang, Y., and J. M. G. Taylor. 2002. A measure of the proportion of treatment effect explained by a surrogate marker. *Biometrics* 58(4):803–812.

Webb, K. S., and G. H. Lin. 1980. Urinary fibronectin—Potential as a biomarker in prostatic-cancer. *Investigative Urology* 17(5):401–404.

Weir, C. J., and R. J. Walley. 2006. Statistical evaluation of biomarkers as surrogate endpoints: A literature review. *Statistics in Medicine* 25(2):183–203.

Wittes, J., E. Lakatos, and J. Probstfield. 1989. Surrogate endpoints in clinical trials: Cardiovascular diseases. *Statistics in Medicine* 8(4):415–425.

Wood, P. H. 1983. Where are we now with radiographic assessment of rheumatoid arthritis? *British Journal of Rheumatology* 22(3 Suppl):24–33.

Woosley, R. L. 1990. CAST—Implications for drug development—Commentary. *Clinical Pharmacology & Therapeutics* 47(5):553–556.

Xu, J., and S. L. Zeger. 2001. The evaluation of multiple surrogate endpoints. *Biometrics* 57(1):81–87.

Appendix B

Recommendations from Related IOM Reports

BOX B-1
Summary of Recommendations to Develop Biomarker-Based Tools for Cancer

Methods, Tools, and Resources Needed to Discover and Develop Biomarkers

1. Federal agencies should develop an organized, comprehensive approach to biomarker discovery, and foster development of novel technologies.
2. Industry and other funders should establish international consortia to generate and share precompetitive data on the validation and qualification of biomarkers.
3. Funders should place a major emphasis on developing quantitative pathway biomarkers to broaden applicability.
4. Funders should sponsor demonstration projects to develop biomarkers that can predict efficacy and safety in patients for drugs already on the market.
5. Government agencies and other funders should sustain support for high-quality biorepositories of prospectively collected samples.

Guidelines, Standards, Oversight, and Incentives Needed for Biomarker Development

6. Government agencies and other stakeholders should develop a transparent process to create well-defined consensus standards and guidelines for biomarker development, validation, qualification, and use.
7. The Food and Drug Administration (FDA) and industry should work together to facilitate the codevelopment and approval of diagnostic–therapeutic combinations.
8. The FDA should clearly delineate and standardize its oversight of biomarker tests used in clinical decision making.
9. The Centers for Medicare & Medicaid Services (CMS) should develop a specialty area for molecular diagnostics under the Clinical Laboratory Improvement Amendments.

Methods and Processes Needed for Clinical Evaluation and Adoption

10. CMS should revise and modernize its coding and pricing system for diagnostic tests.
11. CMS, as well as other payers, should develop criteria for conditional coverage of new biomarker tests.
12. As a component of conditional coverage, establish procedures for high-quality, population-based assessments of efficacy and cost effectiveness of biomarker tests.

SOURCE: IOM (2007a).

BOX B-2
Summary of Recommendations for *The Future of Drug Safety: Promoting and Protecting the Health of the Public*

Organizational Culture

3.1 The committee recommends that the Food, Drug, and Cosmetics Act (FDCA) be amended to require that the Food and Drug Administration (FDA) Commissioner currently appointed by the President with the advice and consent of the Senate also be appointed for a 6-year term of office. The Commissioner should be an individual with appropriate expertise to head a science-based agency, demonstrated capacity to lead and inspire, and a proven commitment to public health, scientific integrity, transparency, and communication. The President may remove the Commissioner from office only for reasons of inefficiency, neglect of duty, or malfeasance in office.

3.2 The committee recommends that an external Management Advisory Board be appointed by the Secretary of Health and Human Services (HHS) to advise the FDA Commissioner in shepherding the Center for Drug Evaluation and Research, or CDER (and the agency as a whole) to implement and sustain the changes necessary to transform the center's culture—by improving morale and retention of professional staff, strengthening transparency, restoring credibility, and creating a culture of safety based upon a lifecycle approach to risk–benefit.

3.3 The committee recommends the Secretary of HHS direct the FDA Commissioner and director of CDER, with the assistance of the Management Advisory Board, to develop a comprehensive strategy for sustained cultural change that positions the agency to fulfill its mission, including protecting the health of the public.

3.4 The committee recommends that CDER appoint an Office of Surveillance and Epidemiology (OSE) staff member to each New Drug Application (NDA) review team and assign joint authority to the Office of New Drugs and OSE for postapproval regulatory actions related to safety.

3.5 To restore appropriate balance between the FDA's dual goals of speeding access to innovative drugs and ensuring drug safety over the product's lifecycle, the committee recommends that Congress should introduce specific safety-related performance goals in the Prescription Drug User Fee Act IV in 2007. (See Chapter 3 for suggested goals.)

Science and Expertise

4.1 The committee recommends that in order to improve the generation of new safety signals and hypotheses, CDER should take the following actions:

BOX B-2 Continued

(a) Conduct a systematic, scientific review of the Adverse Event Reporting System;
(b) Identify and implement changes in key factors that could lead to a more efficient system; and
(c) Systematically implement statistical surveillance methods on a regular and routine basis for the automated generation of new safety signals.

4.2 The committee recommends that in order to facilitate the formulation and testing of drug safety hypotheses, CDER should do the following:

(a) Increase their intramural and extramural programs that access and study data from large automated healthcare databases;
(b) Include in these programs studies on drug use patterns and background incidence rates for adverse events of interest; and
(c) Develop and implement active surveillance of specific drugs and diseases as needed in a variety of settings.

4.3 The committee recommends that the Secretary of HHS, working with the Secretaries of Veterans Affairs and Defense, develop a public–private partnership with drug sponsors, public and private insurers, for-profit and not-for-profit healthcare provider organizations, consumer groups, and large pharmaceutical companies to prioritize, plan, and organize funding for confirmatory drug safety and efficacy studies of public health importance. Congress should capitalize the public share of this partnership.

4.4 The committee recommends that CDER assure the performance of timely and scientifically valid evaluations (whether done internally or by industry sponsors) of Risk Minimization Action Plans (RiskMAPs). The assessment of risks and benefits is an activity that does not end at approval, and risk and benefit cannot be considered in isolation of one another.

4.5 The committee recommends that CDER develop and continually improve a systematic approach to risk–benefit analysis for use throughout the FDA in the preapproval and postapproval settings.

4.6 The committee recommends that CDER build internal epidemiologic and informatics capacity in order to improve the postmarket assessment of drugs.

4.7 The committee recommends that the Commissioner of FDA demonstrate commitment to building the Agency's scientific research capacity by:

(a) Appointing a Chief Scientist in the office of the Commissioner with responsibility for overseeing, coordinating, and ensuring the quality and regulatory focus of the agency's intramural research programs;

BOX B-2 Continued

(b) Designating the FDA's Science Board as the extramural advisory committee to the Chief Scientist;
(c) Including research capacity in the Agency's mission statement;
(d) Applying resources to support intramural research approved by the Chief Scientist; and
(e) Ensuring that adequate funding to support the intramural research program is requested in the Agency's annual budget request to Congress.

4.8 The committee recommends that the FDA have its advisory committees review all NMEs either prior to approval or soon after approval to advise in the process of ensuring drug safety and efficacy or managing drug risks.

4.9 The committee recommends that all FDA drug product advisory committees, and any other peer-review effort such as mentioned above for CDER-reviewed product safety, include a pharmacoepidemiologist or an individual with comparable public health expertise in studying the safety of medical products.

4.10 The committee recommends the FDA establish a requirement that a substantial majority of the members of each advisory committee be free of significant financial involvement with companies whose interests may be affected by the committee's deliberations.

4.11 To ensure that trial registration is mandatory, systematic, standardized, and complete, and that the registration site is able to accommodate the reporting of trial results, the committee recommends that Congress require industry sponsors to register in a timely manner at ClinicalTrials.gov, at a minimum, all phase II through IV clinical trials, wherever they may have been conducted, if data from the trials are intended to be submitted to the FDA as part of an NDA or supplemental NDA, or to fulfill a postmarket commitment. The committee further recommends that this requirement include the posting of a structured field summary of the efficacy and safety results of the studies.

4.12 The committee recommends that the FDA post all NDA review packages on the agency's website.

4.13 The committee recommends that the CDER review teams regularly and systematically analyze all postmarket study results and make public their assessment of the significance of the results with regard to the integration of risk and benefit information.

Regulation

5.1 The committee recommends that Congress ensure that the Food and Drug Administration has the ability to require such postmarketing risk assessment and risk management programs as are needed to monitor and ensure safe use of drug products. These conditions may be imposed both

BOX B-2 Continued

before and after approval of a new drug, new indication, or new dosage, as well as after identification of new contraindications or patterns of adverse events. The limitations imposed should match the specific safety concerns and benefits presented by the drug product. The risk assessment and risk management program may include:

(a) Distribution conditioned on compliance with agency-initiated changes in drug labels;
(b) Distribution conditioned on specific warnings to be incorporated into all promotional materials (including broadcast direct-to-consumer [DTC] advertising);
(c) Distribution conditioned on a moratorium on DTC advertising;
(d) Distribution restricted to certain facilities, pharmacists, or physicians with special training or experience;
(e) Distribution conditioned on the performance of specified medical procedures;
(f) Distribution conditioned on the performance of specified additional clinical trials or other studies; and/or
(g) Distribution conditioned on the maintenance of an active adverse event surveillance system.

5.2 The committee recommends that Congress provide oversight and enact any needed legislation to ensure compliance by both the Food and Drug Administration and drug sponsors with the provisions listed above. The FDA needs increased enforcement authority and better enforcement tools directed at drug sponsors, which should include fines, injunctions, and withdrawal of drug approval.

5.3 The committee recommends that Congress amend the FDCA to require that product labels carry a special symbol such as the black triangle used in the United Kingdom or an equivalent symbol for new drugs, new combinations of active substances, and new systems of delivery of existing drugs. The FDA should restrict DTC advertising during the period of time the special symbol is in effect.

5.4 The committee recommends that the FDA evaluate all new data on new molecular entities no later than 5 years after approval. Sponsors will submit a report of accumulated data relevant to drug safety and efficacy, including any additional data published in a peer-reviewed journal, and will report on the status of any applicable conditions imposed on the distribution of the drug called for at or after the time of approval.

Communication

6.1 The committee recommends that Congress enact legislation establishing a new FDA advisory committee on communication with patients and consumers. The committee would be composed of members who represent consumer and patient perspectives and organizations. The advisory committee would advise CDER and other centers on communication issues

BOX B-2 Continued

related to efficacy, safety, and use during the lifecycle of drugs and other medical products, and it would support the centers in their mission to "help the public get the accurate, science-based information they need to use medicines and foods to improve their health."

6.2 The committee recommends that the new Office of Drug Safety Policy and Communication should develop a cohesive risk communication plan that includes, at a minimum, a review of all center risk communication activities, evaluation and revision of communication tools for clarity and consistency, and priority setting to ensure efficient use of resources.

Resources

7.1 To support improvements in drug safety and efficacy activities over a product's lifecycle, the committee recommends that the Administration should request and Congress should approve substantially increased resources in both funds and personnel for the Food and Drug Administration.

SOURCE: IOM (2007b).

REFERENCES

IOM (Institute of Medicine). 2007a. *Cancer biomarkers: The promises and challenges of improving detection and treatment*. Washington, DC: The National Academies Press.

IOM. 2007b. *The future of drug safety: Promoting and protecting the health of the public*. Washington, DC: The National Academies Press.

Appendix C

Committee Member and Consultant Biographies

John R. Ball, M.D., J.D. (*Chair*), is the executive vice president of the American Society for Clinical Pathology. He is a graduate of Emory University, received a J.D. and an M.D. from Duke University, and was a Robert Wood Johnson Clinical Scholar at George Washington University. After a residency in Internal Medicine at Duke Medical Center, he held several health policy positions within the U.S. Department of Health and Human Services and was a senior policy analyst in the Office of Science and Technology Policy in the Executive Office of the President from 1978 to 1981. From 1986 to 1994, Dr. Ball was the executive vice president of the American College of Physicians, and from 1995 to 1999 he served as president and chief executive officer of the Pennsylvania Hospital. Dr. Ball was the senior vice president and practice director of The Lewin Group's Health Care Organizations Practice in 2000. He is also a member of the American Clinical and Climatological Association and the Society of Medical Administrators.

Michelle A. Albert, M.D., M.P.H., is an assistant professor of medicine at Harvard Medical School and Brigham and Women's Hospital, where she is also an associate physician and the director of behavioral and neurocardiovascular cardiology. Dr. Albert is a board-certified internist and cardiologist, with joint appointments in the Division of Cardiovascular Diseases and the Division of Preventive Medicine, Brigham and Women's Hospital. Dr. Albert earned her M.D. from the University of Rochester School of Medicine and Dentistry, and completed her Internal Medi-

cine internship, residency, and chief residency at Columbia University Presbyterian Medical Center. She then completed a clinical and research fellowship in Cardiovascular Diseases at the Brigham and Women's Hospital. Dr. Albert also completed a research fellowship in Cardiovascular Epidemiology at that hospital and, during that time, earned an M.P.H. with a concentration in Epidemiology from the Harvard School of Public Health.

Dr. Albert devotes approximately 3 months of the year to clinical duties and 8 months to research activities. Specifically, she is an attending physician for the clinical cardiovascular services and cardiac intensive care units at Brigham and Women's Hospital. In these roles, her clinical responsibilities include care for the most critically ill cardiovascular patients. Dr. Albert's research interest concerns the molecular and genetic epidemiology of hemostasis, thrombosis, and inflammation. A central component of her work involves cardiovascular risk prediction using novel biochemical markers in large population-based cohorts. Specifically, Dr. Albert is involved in comparative cardiovascular risk assessment in different race/ethnic groups and international populations as well as work related to the role of chronic psychological stress on cardiovascular outcome. As the Principal Investigator (PI) of a Doris Duke Clinical Scientist Development Award, she piloted a blood collection project to evaluate novel biochemical markers and genetic polymorphisms of cardiovascular risk among geographically diverse Black women living in the United States. Funding from the Donald W. Reynolds Foundation as an Associate Investigator also supports this work. Her research time encompasses developing clinical and epidemiologic cohorts, devising clinical hypotheses, performing data analyses, writing manuscripts, and presenting her findings at numerous venues.

Dr. Albert also teaches Harvard Medical School students and residents at the Veterans Administration (VA) Medical Center and Brigham and Women's Hospital. Dr. Albert is actively involved in several professional organizations, including the Association of Black Cardiologists, American Heart Association (AHA), and the American College of Cardiology (ACC).

Fred Apple, Ph.D., is a professor in the Department of Laboratory Medicine and Pathology at the University of Minnesota–Twin Cities. He also serves as the medical director of Clinical Laboratories and the Clinical Chemistry and Toxicology Laboratories at Hennepin County Medical Center. In addition, he serves as the forensic toxicology consultant to the Hennepin County Medical Examiner's Office. Dr. Apple instructs residents and Fellows during their Chemistry and Toxicology rotations, and directs the Ph.D. Clinical Chemistry Fellowship program. Dr. Apple's

main research interest is the molecular mechanisms that control protein expression in ischemic, necrotic heart, and skeletal muscle. He is also involved in directing and participating in national multicenter clinical trial studies involving biochemical markers of myocardial injury. In addition, Dr. Apple consults with diagnostic, medical instrument, and pharmaceutical industries in the areas of forensic toxicology, clinical toxicology, and chemistry through his limited liability company.

Dr. Apple has been actively involved in National Institutes of Health (NIH), VA, and industry-sponsored research studies involving the application of cardiac, vascular, inflammatory, and ischemic biomarkers for the detection of myocardial cell damage and reperfusion injury identified as either ischemic or inflammatory. He also has an interest in risk stratification and outcomes research. In addition, he has led investigations in sports medicine that concern the biochemistry of exercise. Dr. Apple is also an active participant in studies addressing forensic/postmortem analysis of drugs and alcohol in tissues, blood, and vitreous humor. Other endeavors involve ethanol pharmacokinetics and substance testing. Additionally, Dr. Apple has participated in a number of professional societies, such as the Academy of Clinical Laboratory Physicians and Scientists, National Committee for Clinical Laboratory Standards, American Association for Clinical Chemistry, and American Board of Clinical Chemistry.

Robert M. Califf, M.D., is currently vice chancellor for Clinical Research, director of the Duke Translational Medicine Institute (DTMI), and professor of medicine in the Division of Cardiology at the Duke University Medical Center. Dr. Califf's role as director of the DTMI, which is funded in part by an NIH Clinical and Translational Science Award (CTSA), includes service as cochair of the Principal Investigators Steering Committee of the CTSA. He is also the cochair of the Clinical Trials Transformation Initiative, a public–private partnership focused on improving the clinical trials system.

Dr. Califf graduated from Duke University, summa cum laude and Phi Beta Kappa, and from Duke University Medical School, where he was selected for Alpha Omega Alpha. He performed his internship and residency at the University of California–San Francisco and his fellowship in Cardiology at Duke University. He is board certified in Internal Medicine and Cardiology and is a master of the ACC. For 10 years, Dr. Califf was the founding director of the highly esteemed Duke Clinical Research Institute (DCRI). Dr. Califf led the DCRI in many of the best known clinical trials in cardiovascular disease. With an annual budget of over $100 million, the DCRI has more than 1,000 employees and collaborates extensively with government agencies, the medical products industry, and academic partners around the globe in all therapeutic areas.

Dr. Califf occasionally provides advice to industry on specific projects pertaining to biomarkers.

Additionally, Dr. Califf was the founding director of the coordinating center for the Centers for Education & Research on Therapeutics, a public–private partnership among the Agency for Healthcare Research and Quality, the Food and Drug Administration (FDA), academia, the medical products industry, and consumer groups. This partnership focuses on research and education that will advance the best use of medical products.

Dr. Califf is the editor-in-chief of Elsevier's *American Heart Journal*, the oldest cardiovascular specialty journal. He has been author or coauthor of more than 900 peer-reviewed journal articles and a contributing editor for theheart.org, an online information resource for academic and practicing cardiologists. In cooperation with his colleagues from the Duke Databank for Cardiovascular Disease, Dr. Califf has written extensively about the clinical and economic outcomes of chronic heart disease. He is considered an international leader in the fields of health outcomes, quality of care, and medical economics. He was recently acknowledged as one of the 10 most cited authors in the field of medicine by the Institute for Scientific Information.

Dr. Califf has served on the Cardiorenal Advisory Panel of the FDA and the Pharmaceutical Roundtable of the Institute of Medicine (IOM). He served on the IOM committees that recommended Medicare coverage of clinical trials as well as the removal of ephedra from the market and on the IOM's Committee on Identifying and Preventing Medication Errors. He is currently a member of the IOM Forum on Drug Discovery, Development, and Translation and a subcommittee of the Science Board of the FDA.

Victor G. De Gruttola, Sc.D., is professor and chair of the Department of Biostatistics at the Harvard School of Public Health. Dr. De Gruttola received an M.S. in Bioengineering from Harvard University. He completed his M.S. in Epidemiology and his Doctorate of Science in Biostatistics at the Harvard School of Public Health. Dr. De Gruttola's research activities concern developments of statistical methods required for an appropriate public health response to the AIDS epidemic. His work has been instrumental in outlining the transmission of HIV, natural history of infection with HIV, and clinical research on AIDS therapies.

His work involves not only development and application of statistical methodology for analysis of data from clinical research studies and public health surveillance systems, but also medical issues surrounding HIV infection and concerns of communities most affected by the epidemic. His research goals have included forecasting future AIDS incidence, devel-

oping strategies for clinical research on HIV infection, and evaluating the public health impact of antiviral treatment. A major focus of this research has been on the development and consequences of resistance to antiretroviral drugs. Dr. De Gruttola has also been engaged in statistical issues such as the degree to which the treatment response of HIV-related biomarkers constitute adequate evidence for clinical efficacy. Because of the lack of standard methods to assess the value of data obtained from various biological markers for predicting drug efficacy, he is interested in developing methods to understand and interpret these data. He has also worked on projections of AIDS incidence using data from the New York City Health Department. A special focus of this work was estimation of the risk that children of HIV-infected mothers would develop AIDS in the first 10 years of life using data combined from a variety of sources. Dr. De Gruttola's current interest is in the development of methods for evaluating the plausibility of achieving control of HIV in specific settings using available interventions, and evaluating the community-level impact of pilot studies of combinations of such interventions.

David L. DeMets, Ph.D., is a well-known biomedical researcher and medical educator at the University of Wisconsin–Madison. He is a professor of Statistics and Biostatistics and chair of the Department of Biostatistics and Medical Informatics. Dr. DeMets received his Master's and Doctoral Degrees in Biostatistics from the University of Minnesota. He then completed a postdoctoral appointment within NIH, followed by 10 years at the National Heart, Lung, and Blood Institute (NHLBI), where he eventually assumed the role of chief of the Biostatistics Research Branch. Dr. DeMets developed an interest in the use of statistical methods to design, monitor, and analyze clinical trials. A focus of his work has been in sequential design with specific application to data interim monitoring of accumulating data in an ongoing clinical trial. Dr. DeMets is a Fellow of the International Statistics Institute, the American Statistical Association, the Society for Clinical Trials, the American College of Medical Informatics, and The American Association for the Advancement of Science. He has served on the Board of Directors for the American Statistical Association and Society for Controlled Clinical Trials. He also served terms as president of the Society for Clinical Trials and the Eastern North American Region of the Biometric Society. He has coauthored or coedited four books, titled *Fundamentals of Clinical Trials*, *Data Monitoring Committees in Clinical Trials*, and *Data Monitoring in Clinical Trials: A Case Studies Approach*. Dr. DeMets serves on data safety monitoring boards for industry clinical trials.

Robert Gerszten, M.D., is a senior associate in the Broad Institute of Harvard and the Massachusetts Institute of Technology (MIT) and asso-

ciate professor of medicine in the Harvard Medical School. He is a PI in the Massachusetts General Hospital (MGH) Center for Immunology and Inflammatory Diseases and in the MGH Cardiovascular Research Center.

Dr. Gerszten earned his M.D. Johns Hopkins University and completed his residency at the University of Pennsylvania Hospital. He is board certified in Cardiovascular Disease by the American Board of Internal Medicine. After completion of clinical training in Internal Medicine, Dr. Gerszten was a Postdoctoral Research Fellow in the laboratory of Dr. Shaun Coughlin in the Cardiovascular Research Institute at the University of California–San Francisco. Dr. Gerszten's major focus in the Coughlin laboratory was the structure and function of the human thrombin receptor. Thereafter, he performed cardiac research with Dr. Anthony Rosenzweig to study the molecular mechanisms of leukocyte recruitment as it relates to human inflammatory diseases, with an emphasis on cardiovascular diseases.

Dr. Gerszten's research focuses on the use of unbiased proteomic technologies to identify novel signals in inflammation and wound healing in cardiovascular biology. His overall goal is to identify new metabolites and proteins that mark disease activity, shed insight into disease progression, and ultimately provide targets for therapeutic intervention. His research incorporates basic molecular and cell biology, chemistry and mass spectrometry, and bioinformatics, all with a foundation in clinical medicine. As a member of the Harvard–Partners Proteomics Steering Committee and the leader of a metabolomics effort at the Broad Institute of MIT and Harvard, Dr. Gerszten uses emerging proteomics and metabolomics technologies to help identify novel signals derived from leukocytes, endothelial cells, or the myocardium. In ongoing translational studies, Dr. Gerszten's lab applies these same methodologies directly to samples from well-phenotyped human cohorts to identify candidate biomarkers, which will undergo investigation to determine their functional roles. Dr. Gerszten holds pending patent applications resulting from his biomarker research.

William R. Harlan, Jr., M.D., FACP, FACPM, FAAFP, FAHA, received his M.D. magna cum laude from the Medical College of Virginia and trained in Internal Medicine at Duke University and subsequently had training in Cardiology and Biochemistry at that institution. He has been a professor of medicine at Duke University, University of Alabama, Birmingham, and The University of Michigan. At these last two universities, he was also an Associate Dean within these schools of medicine. In 1987, Dr. Harlan became director of the Division of Epidemiology and Clinical Applications at the NHLBI. In this position, he was responsible for the portfolio

development and oversight of observational and interventional clinical research studies supported by the Institute. In 1991, Dr. Harlan was named associate director for Disease Prevention for NIH and charged with the development of the Women's Health Initiative, a large multidimensional set of clinical trials and observational studies. He was also responsible for the development of the National Center for Complementary and Alternative Medicine and the Office of Dietary Supplements at NIH and served as interim director of each. Dr. Harlan retired from the government in 2001 and has served as a consultant and senior advisor to the Division of Services and Intervention Research at the National Institute of Mental Health and the Office of Dietary Supplements. He has recently served as a senior consultant with the National Library of Medicine on clinical trials registration and a database of results that are part of ClinicalTrials.gov.

Allan S. Jaffe, M.D., is a professor of medicine in the Cardiovascular Diseases Division with a joint appointment in the Department of Laboratory Medicine and Pathology at the Mayo Clinic College of Medicine. He is chair of the Division of Clinical Core Laboratory Services, which involves most of the clinical chemistry operations at the Mayo Clinic. He received his M.D. from the University of Maryland in Baltimore. He was a house officer in Internal Medicine and a Cardiology Fellow at Washington University, and remained on the faculty there from 1978 to 1995. At Washington University, he was instrumental in the clinical development of the cardiac troponin I assays. When he left to become the chief of cardiology and associate chief of medicine at the State University of New York at Syracuse, where he was a full professor with an international academic reputation in the fields of acute ischemic heart disease, depression, heart disease, advanced cardiac life support, and biomarkers of cardiac injury.

He was in Syracuse from 1995 to 1999, when he left to continue his academic career at the Mayo Clinic. He continues to be internationally known for his work in the areas mentioned above. He was cochair of the Enhancing Recovery in Coronary Heart Disease (ENRICHD) trial and chair of the Biochemistry Panel of the European Society of Cardiology and ACC task force for the redefinition of myocardial infarction. Dr. Jaffe was also instrumental in composing the National Academy of Clinical Biochemistry and International Federation of Clinical Chemistry Committee for Standardization of Markers of Cardiac Damage Laboratory Medicine Practice Guidelines: Analytical Issues for Biomarkers of Heart Failure.

He is on seven editorial boards, including *Journal of the American College of Cardiology*, *American Journal of Cardiology*, *Clinical Chemistry*, *Clinical Proteomics*, *Psychosomatic Medicine*, *Cardiology Today*, and *Cardiovascular Therapeutics*, and he recently rotated off the *Circulation Board*.

Ronald M. Krauss, M.D., is senior scientist and director of Atherosclerosis Research at Children's Hospital Oakland Research Institute, adjunct professor in the Department of Medicine at the University of California, San Francisco and in the Department of Nutritional Sciences at the University of California, Berkeley, and guest senior scientist in the Department of Genome Sciences of Lawrence Berkeley National Laboratory. He received his undergraduate and medical degrees from Harvard University with honors and served his internship and residency on the Harvard Medical Service of Boston City Hospital. He then joined the staff of the National Heart, Lung, and Blood Institute in Bethesda, Maryland, first as clinical associate and then as senior investigator in the Molecular Disease Branch. Dr. Krauss is board-certified in internal medicine, endocrinology and metabolism, and is a member of the American Society for Clinical Investigation, a Fellow of the American Society of Nutrition and the American Heart Association (AHA), and a Distinguished Fellow of the International Atherosclerosis Society. He a member of the U.S. National Cholesterol Education Program Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults, founder and past chair of the AHA Council on Nutrition, Physical Activity, and Metabolism, and a National Spokesperson for the AHA. Dr. Krauss has also served on both the Committee on Dietary Recommended Intakes for Macronutrients and the Committee on Biomarkers in Chronic Disease of the Institute of Medicine of the National Academy of Sciences. He has received numerous awards including the AHA Scientific Councils Distinguished Achievement Award and the Centrum Center For Nutrition Science Award of the American Society for Nutrition, and he is listed in Who's Who in America and the World. Dr. Krauss is on the editorial boards of a number of journals, and is associate editor of *Obesity.* Dr. Krauss has published nearly 400 research articles and reviews on genetic, dietary, and drug effects on plasma lipoproteins and coronary artery disease. In recent years Dr. Krauss' work has focused on interactions of genes with dietary and drug treatments that affect metabolic phenotypes and cardiovascular disease risk.

Harlan M. Krumholz, M.D., S.M., is the Harold H. Hines, Jr., Professor of Medicine and Epidemiology and Public Health at Yale University School of Medicine, where he is director of the Robert Wood Johnson Clinical Scholars Program. He is also the director of the Yale–New Haven Hospital Center for Outcomes Research and Evaluation. He received his M.D. from Harvard Medical School and a Master of Science degree in Health Policy and Management at the Harvard School of Public Health. He trained in Internal Medicine at the University of California–San Francisco and in Cardiology at Beth Israel in Boston.

Dr. Krumholz's research is focused on determining optimal clinical strategies and identifying opportunities for improvement in the prevention, treatment, and outcome of cardiovascular disease, with an emphasis on underrepresented populations. Using methods of clinical epidemiology and health services research, he has sought to illuminate the balance of risks, benefits, and costs of specific clinical approaches. The research efforts are intended to provide critical information to improve the quality of health care, monitor changes over time, and guide decisions about the allocation of scarce resources.

Dr. Krumholz is currently leading initiatives through the Centers for Medicare & Medicaid Services to develop national mortality measures for public reporting of hospital performance. In an effort to investigate ways that hospitals can improve outcomes through decreasing door-to-balloon times, he initiated and chairs the steering committee of D2B: An Alliance for Quality, an international campaign launched by the ACC to implement key evidence-based strategies to achieve guideline-recommended door-to-balloon time. He also serves as PI on two multicenter projects sponsored by the NHLBI: (1) the VIRGO study, an investigation of issues surrounding the care and outcomes of young women with acute myocardial infarction; and (2) a study examining the effect of a telemonitoring strategy on the outcomes of patients with heart failure. Dr. Krumholz is a member of the Association of American Physicians, the American Society for Clinical Investigation, and the IOM. He is also the author of the book *The Expert Guide to Beating Heart Disease*.

Maria Lopes-Virella, M.D., Ph.D., is a professor in the Department of Medicine and the Division of Endocrinology, Diabetes, and Medical Genetics at the Medical University of South Carolina (MUSC). Dr. Lopes-Virella serves as a staff physician, director of the Nutrition Support Team, and chief of the Clinical Chemistry Section of Laboratory Services at the Ralph H. Johnson VA Medical Center at MUSC. Additionally, she has a joint appointment in the Department of Immunology & Microbiology and in the Department of Pathology and Laboratory Medicine, College of Medicine. Dr. Lopes-Virella also has appointments in the Oral and Community Health Sciences Division, Department of Stomatology, College of Dental Medicine at MUSC and in the Department of Bioengineering at Clemson University. She received her M.D. and doctorate degree in medicine/biochemistry from the University of Lisbon, Portugal. Thereafter, she completed Internal Medicine and Clinical Pathology residencies at University Hospital in Lisbon and two fellowships (Endocrinology and Pathology) at MUSC.

Her research interests are modified LDLs as biomarkers for the progression of type 1 diabetes, quantity of iron stores as indicators of acute

coronary heart disease, and the role of matrix metalloproteinases in acute coronary heart disease. Ongoing research includes a recently completed NIH/NHLBI-funded study, "Markers and Mechanisms of Macrovascular Disease in Diabetes Mellitus," in which she studied the role of immune responses to modified lipoproteins and of inflammation in the development of micro- and macrovascular complications in large cohorts of type 1 and type 2 diabetes; a VA Merit Review-funded research project on "Lipoprotein Metabolism and Cell-Lipoprotein Interactions in Diabetes Mellitus"; and a Juvenile Diabetes Foundation International-funded grant to study biomarkers and mechanisms of nephropathy in type 1 diabetes.

Dr. Lopes-Virella is a Fellow of the AHA, American College of Nutrition, American College of Pathology, and Council of Nutrition, Physical Activity, and Metabolism, AHA as well as the Council on Arteriosclerosis, Thrombosis, and Vascular Biology, AHA. She serves on the boards of the National Lipid Association and American Board of Clinical Lipidology. Dr. Lopes-Virella is or has been a distinguished member of a number of professional organizations, including the Southern Society for Clinical Investigation, Endocrine Society, American Diabetes Association, and American Federation of Medical Research. She is the recipient of numerous awards, including a Special Emphasis Research Career Award from the NHLBI and the National Institute of Arthritis Metabolism and Digestive Diseases and a VA Clinical Investigator Award. In 2006, she received the Vision 7 Henry Middleton Award for Excellence in Research. She was the recipient of the Edwin Berman Lectureship Award from the American Diabetes Association, John A. Colwell Award for Outstanding Contributions to Diabetes Research from the South Carolina Affiliate of the American Diabetes Association, and the Mary Jane Kugel Award from the Juvenile Diabetes Foundation International. Dr. Lopes-Virella has served on editorial boards and peer reviews for several journals and recently became an associate editor for the *Journal of Clinical Lipidology.*

Roberta B. Ness, M.D., M.P.H., is the dean of the University of Texas School of Public Health, M. David Low Chair in Public Health, and co-PI of the Center for Clinical and Translational Science at the University of Texas–Houston. Dr. Ness received her M.D. from Cornell University and her M.P.H., with a concentration in Epidemiology, from Columbia University. She completed her Internal Medicine internship and residency at Bellevue Hospital in New York City. While completing her master's program, she participated in the NIH Training Fellowship. Her research interests are in reproductive biology and women's health. Dr. Ness has received acclaim from her colleagues concerning her research that examined the disease origins of pelvic inflammatory disease, preeclampsia, and ovarian cancer. Throughout her career, she investigated biological markers specific

to reproductive dysfunction. Dr. Ness published papers concerning the use of fetal fibronectin to predict ectopic and intrauterine pregnancies, endometriosis markers, and ovarian cancer biological markers. Since her formal training, she has lent her expertise to a number of organizations; she is current president of the American College of Epidemiology. She was a core member of the Early Markers of Adult Disease Workgroup and Study Assembly and codirector for the Symposium on Ovarian Cancer and High-Risk Women: Implications for Prevention, Screening, and Early Detection. She currently gives lectures concerning markers and molecular epidemiology.

Jennifer Van Eyk, Ph.D., is a professor of medicine in the Division of Cardiology and crossappointed to Biological Chemistry and Biomedical Engineering at Johns Hopkins University. She earned her doctoral degree from the University of Alberta. Her research combines physiology and proteomics to provide an in-depth analysis of the molecular basis for a variety of cardiac diseases ranging from myocardial ischemia to heart failure. In addition, her group develops serum/plasma biomarkers in which de novo discovery is coupled with validation. Dr. Van Eyk holds patents resulting from her biomarker research.

Dr. Van Eyk has been a Canadian Heart and Stroke Foundation Scholar (1996–2001) and Heart and Stroke Career Investigator (received in 2001), in addition to receiving a Canadian Institutes of Health Research Investigator Award prior to being recruited to Johns Hopkins. Dr. Van Eyk is a Fellow of the AHA, current chair of the Genomics and Translational Science Council for the AHA, a member of the senior editorial board of *Proteomics: Clinical Application*, and on the editorial board of the *Journal of Physiology*. She was a guest editor for a series on proteomics in a number of journals and has coedited two books in this area: *Proteomic and Genomic Analysis of Cardiovascular Disease* and *Clinical Proteomics: From Diagnosis to Therapy*.

John A. Wagner, M.D., Ph.D., is vice president of Clinical Pharmacology and Acting Modeling and Simulation Integrator in Strategically Integrated Modeling and Simulation at Merck & Co. Dr. Wagner is also an adjunct assistant professor in the Division of Clinical Pharmacology within the Department of Medicine at Jefferson Medical College at Thomas Jefferson University. He is a visiting clinical scientist within the Harvard–MIT Division of Health Sciences and Technology in the Center for Experimental Pharmacology and Therapeutics. He received his M.D. from Stanford University School of Medicine and his Ph.D. from Johns Hopkins University School of Medicine. His postgraduate training is in Internal Medicine and Molecular and Clinical Pharmacology. He is chair of the Pharma-

ceutical Research and Manufacturers of America Clinical Pharmacology Technical Group, chair of the adiponectin workgroup for the Biomarkers Consortium, and member of the board of directors for the American Society for Clinical Pharmacology and Therapeutics. His publications are in the areas of biomarkers and surrogate endpoints, experimental medicine, pharmacokinetics, pharmacodynamics, and drug–drug interactions across a variety of therapeutic disciplines. Dr. Wagner was coauthor on an influential perspective in *Nature Reviews Drug Discovery*, entitled "A cost-effectiveness approach to the qualification and acceptance of biomarkers" and a paper in *Clinical Pharmacology and Therapeutics* on "Biomarkers and surrogate end points for fit-for-purpose development and regulatory evaluation of new drugs."

Elizabeth A. Yetley, Ph.D., who served as the consultant to the committee, is a retired government scientist. Her career spans more than 28 years of government service, including 24 years at the FDA, culminating with her appointment as lead scientist for nutrition. From 2004 until her retirement, she was a senior nutrition research scientist with the NIH Office of Dietary Supplements. Her leadership activities in the field of nutrition public health policy have had a considerable impact. She has been responsible for nutrition labeling (particularly for the development of the health claims paradigm), national food fortification programs, use of national nutrition monitoring and surveillance systems to support nutrition and food safety health policies, infant formula and medical food reviews and regulatory oversight, dietary supplement regulation, and the use of nutrient-related reference values in public health policy formulation. She has received more than 75 honors, commendations, and letters of recognition for her service and has served as a scientific representative for the government to more than 50 associations, panels, and committees. She has authored or coauthored approximately 100 scientific and peer-reviewed publications. Dr. Yetley received her Ph.D. in nutrition, with a minor in biochemistry and physiology, from Iowa State University.

Appendix D

Staff Biographies

Christine M. Micheel, Ph.D., joined the Institute of Medicine (IOM) in 2008 as a Mirzayan Science and Technology Graduate Fellow in the National Cancer Policy Forum (NCPF). She has worked on NCPF activities including "Improving the Quality of Cancer Clinical Trials," "Implementing Colorectal Cancer Screening," and "Multi-site Phase III Clinical Trials and NCI [National Cancer Institute] Cooperative Groups." She is now the study director for the Committee on Qualification of Biomarkers and Surrogate Endpoints in Chronic Disease, which is housed in the Board on Health Care Services. Prior to joining the IOM, Dr. Micheel completed a postdoctoral position at the IBM Almaden Research Center in San Jose, CA, where she studied interactions between biomolecules—such as DNA and antibodies—and nanomaterials. She completed her Ph.D. in Chemistry at the University of California–Berkeley in 2005, under the direction of Paul Alivisatos and with the support of a Howard Hughes Medical Institute Predoctoral Fellowship. Her research was focused at the boundary between nanoscience and biophysics. Outside of her research pursuits, Dr. Micheel volunteered in the library at the Women's Cancer Resource Center in Oakland, CA, a community resource for women with cancer and their families. Dr. Micheel obtained her Bachelor's Degree at Washington University in St. Louis, MO, where she graduated magna cum laude with a major in Chemistry. Her undergraduate studies were supported by a Compton II fellowship (now known as the Florence Moog Fellowship) for pursuit of studies in the fields of biology and chemistry.

Sharyl Nass, Ph.D., is a study director and senior program officer at the IOM, where she has worked with the Board on Health Sciences Policy, the Board on Health Care Services, and the National Cancer Policy Board and Forum. She was also recently named as the director of the IOM's National Cancer Policy Forum. Her previous work at the IOM has focused on topics that include developing cancer biomarkers, formulating strategies for large-scale biomedical science, developing technologies for the early detection of breast cancer, improving breast imaging quality standards, improving the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, and facilitating contraceptive research and development. Her current IOM position combines her dual interests in biomedical research and health science policy. With a Ph.D. in Cell and Tumor Biology from Georgetown University and postdoctoral training at the Johns Hopkins University School of Medicine, she has authored numerous papers on the cell and molecular biology of breast cancer. She also holds a B.S. in Genetics and an M.S. in Endocrinology/Reproductive Physiology, both from the University of Wisconsin–Madison. In addition, she studied developmental genetics and molecular biology at the Max Planck Institute in Germany under a fellowship from Fulbright and the German Heinrich Hertz-Stiftung Foundation. Dr. Nass was the 2007 recipient of the IOM's Cecil Award for Excellence in Health Policy Research.

Erin Balogh, M.P.H., joined the IOM in 2008 as a research associate for the NCPF and Board on Health Care Services. She is currently working on two committee studies, the Qualification of Biomarkers and Surrogate Endpoints in Chronic Disease and Cancer Clinical Trials and the NCI Cooperative Group Program. She completed her M.P.H. at the University of Michigan in Health Management and Policy, and prior to that, graduated summa cum laude from Arizona State University with her Bachelor's Degrees in Microbiology and Psychology. Ms. Balogh interned with AcademyHealth in Washington, DC, and worked as a research site coordinator for the Urban Institute in Topeka, Kansas. As an undergraduate, Ms. Balogh worked as a management intern with the Arizona State University Office of University Initiatives, a strategic planning group for the university.

Bernadette McFadden, M.Sc., joined the IOM as a research associate in 2008. Since that time, she has staffed projects on the redesign of continuing education for health professionals and the standardization of race, ethnicity, and language data. She currently works with the Committee on Qualification of Biomarkers and Surrogate Endpoints in Chronic Disease and the Committee on Future Directions for the National Healthcare

Quality and Disparities Reports. Prior to joining the IOM, she completed a Master's Degree in Social Research at Trinity College Dublin and was employed by Dublin City Council's Homeless Agency, where she edited a volume of essays on homelessness in Ireland and wrote a report on how the city's management of public space impacts homeless persons. She graduated summa cum laude, Phi Beta Kappa, from Dickinson College in Pennsylvania. While in Pennsylvania, she conducted research on local effects of implementing Medicare Part D and the state's long-term care policies; interned with the Executive Policy Office of the Pennsylvania Department of Health; and served as a board member for the United Way of Cumberland County. Her interests in health policy developed while serving as an AmeriCorps teacher in an Atlanta public school.

Lisa Boyette, M.D., completed her M.D. at the University of Virginia in 2007 and is now working on a Ph.D. in Molecular Physiology and Biological Physics at the National Institutes of Health (NIH). Her research at NIH focuses on stem cell reprogramming techniques and how reprogramming technology can be applied to cell-based therapies and tissue engineering. Dr. Boyette studied biomedical engineering and physics as an undergraduate at Virginia Commonwealth University and the Medical College of Virginia. Following completion of her Ph.D., she plans to complete residency training in Neurosurgery. Through her Mirzayan Fellowship with the National Cancer Policy Forum, she learned about crafting policy that promotes research that will effectively advance the standard of care provided to patients.

Anna Woloszynska-Read, Ph.D., was a Christine Mirzayan Science and Technology Policy Graduate Fellow with the NCPF from January to April 2009. During her Fellowship, Dr. Woloszynska-Read contributed to the study on "Qualification of Biomarkers and Surrogate Endpoints in Chronic Disease" and the study on "Cancer Clinical Trials and the NCI Cooperative Groups." She completed her Ph.D. in Molecular Pharmacology at Roswell Park Cancer Institute in Buffalo, NY, in 2009, with the support of an NIH predoctoral training grant. She holds an M.S. in Human Biology from Adam Mickiewicz University in Poland. Her dissertation work involved translational research relating to the epigenetics of ovarian cancer, with the hopes of developing early diagnostic tools and novel treatments for the disease. During her time as a graduate student, she worked with the Scientific Review Committee and the Institutional Review Board at Roswell Park Cancer Institute. This experience has made her aware of the importance of science policy and the institutional oversight of scientific research. She is currently in a postdoctoral research position at Roswell Park Cancer Institute that combines her interests in

basic science and policy, focusing on translational research and issues of health disparities.

Caira M. Woods, Ph.D., is a Christine Mirzayan Science and Technology Policy Fellow with the Board on Health Care Services and National Cancer Policy Forum. Her interests center around federal-level health and science policy. She completed her Ph.D. in basic medical science at New York University in December 2009. Her research was funded by a National Research Service Award from the National Institutes of Health (NIH) and received an honorable mention from the National Science Foundation (NSF) Graduate Research Fellowship Program. Dr. Woods is a member of Phi Beta Kappa National Honor Society and a magna cum laude graduate of Spelman College in Atlanta, Georgia. During and after college, she participated in summer science programs at Spelman, the NIH, Duke University Medical Center and the NSF. As an intern, Dr. Woods evaluated NSF-funded workshops on broadening participation. This experience introduced her to science policy and solidified her desire to pursue a career in this field. Dr. Woods is particularly interested in public understanding of science, broadening participation, global competitiveness and health policy and is thrilled to have the opportunity to gain a more in-depth perspective on the role scientists can play in implementing and improving programs in these areas.

Ashley McWilliams joined the IOM in September 2008 as a senior program assistant for the Board on Health Care Services and the National Cancer Policy Forum. At the IOM, she is working on projects such as the Committee on Qualification of Biomarkers and Surrogate Endpoints in Chronic Disease and the Workshop Planning Committee for the National Emergency Care Enterprise. She has also worked with the IOM's Roundtable on Evidence-based Medicine and the Office of Reports and Communication. Prior to joining the IOM, Ms. McWilliams graduated magna cum laude and Phi Beta Kappa from Howard University with a Bachelor's Degree in Biology in 2008. During college, Ms. McWilliams was copresident of the Health Professions Society and a member of several honor societies. Ms. McWilliams has also participated in summer research programs at the University of California–San Francisco, Massachusetts Institute of Technology, and Virginia Polytechnic Institute and State University; she also participated in a summer health careers program at Case Western Reserve University.

Roger Herdman, M.D., is director of the IOM Board on Health Care Services. He received his undergraduate and medical school degrees from Yale University. Following an internship at the University of Minnesota

and a stint in the U.S. Navy, he returned to Minnesota, where he completed a residency in Pediatrics and a Fellowship in Immunology and Nephrology and also served on the faculty. He served as professor of Pediatrics at Albany Medical College until 1979. In 1969, Dr. Herdman was appointed director of the New York State Kidney Disease Institute in Albany, NY, and shortly thereafter was appointed deputy commissioner of the New York State Department of Health (1969–1977). In 1977 he was named New York State's director of public health. From 1979 until joining the U.S. Congress Office of Technology Assessment (OTA), he served as a vice president of Memorial Sloan-Kettering Cancer Center in New York City. In 1983, Dr. Herdman was named assistant director of OTA, where he subsequently served as director from 1993 to 1996. He later joined the IOM as a senior scholar and directed studies on graduate medical education, organ transplantation, silicone breast implants, and the Veterans Administration national formulary. Dr. Herdman was appointed director of the IOM/National Research Council National Cancer Policy Board from 2000 through 2005. From 2005 until 2009, Dr. Herdman directed the IOM National Cancer Policy Forum. In 2007, he was also appointed director of the IOM Board on Health Care Services. During his work at the IOM, Dr. Herdman has worked closely with the U.S. Congress on a wide variety of healthcare policy issues.

Linda D. Meyers, Ph.D., is director of the Food and Nutrition Board (FNB) at the IOM. She is responsible for a portfolio that includes nutrient requirements (Dietary Reference Intakes), obesity prevention, food safety, and international, military, and specific population nutrition. She also directed the FNB's international nutrition program from 1982 to 1986. From 1986 to 2001, she served in the Office of Disease Prevention and Health Promotion in the U.S. Department of Health and Human Services, where she was a senior nutrition advisor, deputy director, and acting director. While there, she oversaw preparation of a number of technical and policy reports, including the 1990, 1995, and 2000 *Dietary Guidelines for Americans*, the *U.S. Action Plan on Food Security*, and the national health objectives for 2010. Dr. Meyers has a B.A. in Health and Physical Education from Goshen College in Indiana, an M.S. in Food and Nutrition from Colorado State University, and a Ph.D. in Nutritional Sciences from Cornell University. Her research has focused on population indicators of nutritional status. She has also worked in Botswana, Kenya, and Vietnam. Dr. Meyers has received a number of awards for her contributions to public health, including the Secretary's Distinguished Service Award for *Healthy People 2010* and the Surgeon General's Medallion.

Andrew Pope, Ph.D., is director of the Board on Health Sciences Policy at the IOM. He has a Ph.D. in Physiology and Biochemistry from the University of Maryland and has been a member of The National Academies staff since 1982 and of the IOM staff since 1989. His primary interests are science policy, biomedical ethics, and environmental and occupational influences on human health. During his tenure at The National Academies, Dr. Pope has directed numerous studies on topics that range from injury control, disability prevention, and biologic markers to the protection of human subjects of research, NIH priority-setting processes, organ procurement and transplantation policy, and the role of science and technology in countering terrorism. Dr. Pope is the recipient of the IOM's Cecil Award and the National Academy of Sciences President's Special Achievement Award.

Appendix E

Workshop Agenda

Institute of Medicine
Committee on Qualification of Biomarkers and Surrogate Endpoints in Chronic Disease, Meeting 2 Workshop

The Keck Center of The National Academies
500 Fifth Street, NW
Washington, DC 20001
April 6, 2009

	OPEN SESSION
8:00 am	Breakfast
8:30 am	Welcome and Opening Remarks *John Ball, Committee Chair*
9:00 am	Analysis of Task—General Guidance—Robert Temple
10:00 am	**BREAK**
10:15 am	Quantitative Decision Analytical Modeling Tools—Rebecca Miksad
11:15 am	Existing Frameworks for Biomarker Qualification—CDER—Marc Walton and Aliza Thompson
12:15 pm	**BREAK**
12:30 pm	**LUNCH**—Analogous Environment—Process Standards for Manufacturing, Businesses—David Dilts

1:30 pm	Existing Frameworks for Biomarker Qualification —PhRMA—James Mayne
2:00 pm	NIH Biomarker Qualification/Cancer Perspective —Arthur Schatzkin
2:30 pm	Risk Factor to Surrogate Endpoint Pathway—Philip Greenland
3:00 pm	Discussion
3:30 pm	**BREAK**
3:45 pm	Troponin—James de Lemos
4:15 pm	CRP and Inflammatory Markers—Christie Ballantyne
4:45 pm	HDL/LDL—Bryan Brewer
5:15 pm	Case Studies Discussion *Speakers and Committee*
5:45 pm	**ADJOURN**

Appendix F

Speaker Biographies

Christie M. Ballantyne, M.D., is director of the Center for Cardiovascular Disease Prevention, Methodist DeBakey Heart Center; chief of the Section of Atherosclerosis and Vascular Medicine, interim chief, Section of Cardiology, Department of Medicine, Baylor College of Medicine; director of the Maria and Alando J. Ballantyne, M.D., Atherosclerosis Laboratory; professor of medicine with a joint appointment in Pediatrics, Baylor College of Medicine; and codirector, Lipid Metabolism and Atherosclerosis Clinic, The Methodist Hospital, Houston, TX. He received his Doctor of Medicine from Baylor College of Medicine, and his postgraduate training included an internal medicine residency at The University of Texas Southwestern Medical School, a cardiology fellowship at Baylor College of Medicine, and an American Heart Association (AHA)/Bugher Foundation Fellowship at the Howard Hughes Medical Institute and Institute for Molecular Genetics at Baylor. Dr. Ballantyne is a Fellow of the American Association for the Advancement of Science, member of the American Society for Clinical Investigation, Fellow of the American College of Cardiology (ACC), and Fellow of the American College of Physicians. He previously served as governor of the Texas Chapter of the ACC and president of the Houston Chapter of the AHA. Dr. Ballantyne has been the recipient of numerous study grants, including an AHA Established Investigator Award and several National Institutes of Health (NIH) grants to study leukocyte–endothelial adhesion molecules and novel biomarkers for atherosclerosis. He has been a member of numerous steering committees for multicenter trials, including the Atherosclerosis

Risk in Communities (ARIC) study, Improved Reduction of Outcomes: Vytorin Efficacy International Trial (IMPROVE IT), A Study to Evaluate the Effect of Rosuvastatin on Intravascular Ultrasound-Derived Coronary Atheroma Burden (ASTEROID), National Cholesterol Education Program Evaluation Project Utilizing Novel E-Technology II (NEPTUNE II), and Effect of Niacin ER/Lovastatin on Peak Walking Time and Claudication Onset Time in Patients with Intermittent Claudication (ICPOP). He has also participated as a member of several data and safety monitoring boards and is editorial director for www.lipidsonline.org. He has published extensively and has spoken nationally and internationally on lipids, atherosclerosis, and inflammation. Dr. Ballantyne's research interests include the pathophysiology of atherosclerosis, with an emphasis on monocyte activation and adhesion. His clinical interests include preventive cardiology, lipids, metabolic syndrome, atherosclerosis, genetics, and coronary artery disease.

Joseph Bonventre, M.D., Ph.D., received his M.D. and Ph.D. (Biophysics) from Harvard University. Dr. Bonventre is the Robert H. Ebert Professor of Medicine and Health Sciences and Technology at Harvard Medical School. He is director of the Renal Division at Brigham and Women's Hospital. His research involves investigating the mechanisms of cellular and tissue injury and repair, particularly as applied to ischemic injury to the kidney.

H. Bryan Brewer, Jr., M.D., is the director of Lipoprotein and Atherosclerosis Research at the Cardiovascular Research Institute at Washington Hospital Center in Washington, DC. He was formerly chief of the Molecular Disease Branch at the National Heart, Lung, and Blood Institute (NHLBI) of the NIH, a position he held from 1976 until 2005. Dr. Brewer's research led to the elucidation of the first published sequences for the human plasma apolipoproteins, the initial determination of the metabolism of the plasma apolipoproteins in normal and hyperlipidemic individuals, as well as the identification of multiple gene defects leading to the genetic dyslipoproteinemias. More recently, he has pioneered the use of transgenic mice and rabbits, as well as recombinant adenovirus vectors to identify genes that modulate lipoprotein metabolism and the development of atherosclerosis. Dr. Brewer received his M.D. from Stanford University School of Medicine. After completing his internship and residency in Internal Medicine at Massachusetts General Hospital, he joined NHLBI. He served as a member of the Board of the National Cholesterol Education Program, which established treatment guidelines for patients with hyperlipidemia in the United States. As a recipient of the J.D. Lane Investigator Award from the U.S. Public Health Service,

Dr. Brewer also received the Heinrich Wieland Prize from the Federal Republic of Germany; the Public Health Service Commendation, Meritorious Service, and Distinguished Service Medals from the NIH; the George Lyman Duff Memorial Commendation Award Lecture; and the Robert I. Levy Award. Dr. Brewer has published more than 450 original manuscripts and 75 reviews and book chapters on the subjects of genetic dyslipoproteinemias, lipoprotein metabolism, and atherosclerosis. He has served on the boards of several prestigious journals and is currently on the editorial board of the *Journal of Biological Chemistry*.

James de Lemos, M.D., is the director of the Coronary Care Unit at Parkland Memorial Hospital and the director of the Cardiology Fellowship at the University of Texas Southwestern Medical School in Dallas, Texas. He is an associate professor of medicine and holds the J. Fred Schoelkopf Endowed Chair in Cardiology. He is an active investigator in the Donald W. Reynolds Cardiovascular Research Center at University of Texas Southwestern, and remains closely affiliated with the Thrombolysis in Myocardial Ischemia research group. His primary research interest is risk assessment and management of acute and chronic coronary artery disease. His other research interests include electrocardiography as a means of assessing the coronary microcirculation after thrombolysis or percutaneous coronary intervention, and the use of novel biomarkers for prognostic assessment among patients with coronary artery disease. He has worked extensively with novel biomarkers such as B-type natriuretic peptide, Monocyte Chemoattractant protein-1, and soluble CD40 ligand. He was recently the lead author of the Z phase of the A to Z trial, an international trial investigating different cholesterol-lowering strategies in patients with acute coronary syndromes. He graduated from Harvard Medical School and completed an Internal Medicine Residency at the University of Texas Southwestern Medical Center, where he also served as chief medical resident. He completed a Fellowship in Cardiovascular Medicine at Brigham and Women's Hospital, and served on the faculty of Brigham and Women's Hospital and Harvard Medical School before moving to University of Texas Southwestern Medical School. He has served on multiple committees of the AHA and ACC and is on the editorial board of the *American Journal of Cardiology* and the *American Heart Journal*. He has authored or coauthored over 150 manuscripts or book chapters and won several teaching awards.

David Dilts, Ph.D., M.B.A., is director of Clinical Research for the Knight Cancer Institute and professor of Healthcare Management at the Oregon Health & Science University. Currently, he is on leave as a joint professor of Management and Engineering Management in the Owen Graduate

School of Management and the Vanderbilt University School of Engineering. He is director of Engineering Management Program and is codirector of the Center for Management Research in Healthcare. This center, funded in part by the National Cancer Institute (NCI), WebMD, and others, has as its primary mission the exchange of knowledge between management and health care to dramatically impact the practice of medicine. One research stream, funded by the NCI, is to apply management principles to significantly reduce the time and steps required to open oncology clinical trials. This research has completed in-depth examinations of four NCI-designated Comprehensive Cancer Centers, two major oncology cooperative groups, the NCI Cancer Therapy Evaluation Program, and the NCI Centralized Institutional Review Board. His work has been published in more than 160 articles, conference papers, presentations, book chapters, books, and monographs. This research has been supported by grants totaling nearly $10 million in the past decade.

Philip Greenland, M.D., is director of the Northwestern University Clinical and Translational Sciences Institute and Principal Investigator of Northwestern University's NIH-funded Clinical and Translational Science Award. He holds the Harry W. Dingman Professorship (Endowed Chair) in Preventive Medicine and was chair of the Department of Preventive Medicine at Northwestern University from 1991 to 2005. In 2005, he was appointed senior associate dean for Clinical and Translational Research at Northwestern's Feinberg School of Medicine. Dr. Greenland holds a B.A. from Williams College in Massachusetts and an M.D. from the University of Rochester School of Medicine and Dentistry. Following postgraduate education in Internal Medicine and Cardiology, Dr. Greenland was an assistant and associate professor in the Departments of Medicine, Preventive and Community Medicine, and Psychiatry at the University of Rochester from 1980 to 1991. In 1989–1990, Dr. Greenland was visiting professor of Cardiology at the Henry Neufeld Cardiovascular Institute at Tel-Hashomer Hospital, Tel-Aviv University, Israel. In 1999–2000, he served as visiting professor in the Department of Preventive Medicine at the Brigham and Women's Hospital, Harvard Medical School, Boston. From 2004–08, Dr. Greenland was editor of the *Archives of Internal Medicine*. He has also served on the editorial boards of *Journal of the American Medical Association (JAMA)*, *American Journal of Epidemiology*, and *Journal of Cardiovascular and Pulmonary Rehabilitation*. He is a current member of the scientific advisory board of *Science—Translational Medicine*. Dr. Greenland's research work is notable in three primary areas. He and his colleagues have demonstrated the substantial role of traditional cardiovascular risk factors in long-term cardiovascular risk assessment. His *JAMA* paper in 2003 has been recognized as an "ISI Classic," a highly

cited paper in clinical medicine, for its key role in addressing the long-held myth that only 50 percent of patients with cardiovascular disease have traditional risk factors present. Other papers by Greenland and colleagues have demonstrated the long-lasting and overwhelming effect of traditional cardiovascular risk factors for long-term risk prediction. Greenland's research was also among the first to show that cardiovascular risks after myocardial infarction are different in women and men. His 1991 paper, cited more than 300 times, inaugurated the field of research on heart disease outcomes in women. He is also regarded as a leader in the selective use of cardiovascular imaging, especially coronary calcium measurement by cardiac computed tomography, in global cardiovascular risk assessment. His 2004 *JAMA* paper on this topic has been recognized as the leading paper in this area, cited more than 200 times. He has chaired or cochaired multiple guidelines panels dealing with risk assessment in cardiovascular disease. Dr. Greenland is a member of the NHLBI Board of External Experts and a member of the NHLBI Monitoring Board for the Framingham Heart Study. He previously served on the NIH Study Section on Cardiovascular and Sleep Epidemiology. He currently chairs the ACC–AHA Guidelines Panel on Assessment of Cardiovascular Risk in the Asymptomatic Individual, and he is a member of the NHLBI Cardiovascular Prevention Guidelines Panel. Dr. Greenland has served as a reviewer of several Institute of Medicine reports, most recently the report on medical effects of the Gulf War.

James T. Mayne, Ph.D., DABT, has more than 20 years of pharmaceutical industry experience, including scientific and managerial leadership roles in Drug Safety and Regulatory Affairs for Pfizer, Inc. Currently, he is the senior director of Regulatory Strategy and Policy for Pfizer's Global Research & Development. In addition to contributions at the project and portfolio levels, he was instrumental in the development and implementation of genomics, proteomics, and metabonomics investigative and biomarker capabilities at Pfizer. More recently Dr. Mayne has worked both internal and external to Pfizer on regulatory strategy and regulatory policy issues, including biomarker development and qualification, drug-induced liver injury, and biotherapeutics development. He has published over 25 original research and review articles on topics related to biomarkers, drug safety, and drug development, and holds of three U.S. and international patents. Dr. Mayne received his Ph.D. and postgraduate training in Comparative Toxicology at Cornell University. He is a past guest lecturer in Safety Risk Assessment and Risk Management at Northeastern University and Yale University, and is currently on the faculty of the Harvard–Massachusetts Institute of Technology Clinical Investigator Training Program, where he provides practical insight into safety assess-

ment and drug development strategies in pharmaceutical research and development. Professional activities include membership in the Drug Information Association, Society of Toxicology, and New York Academy of Sciences, and certification by the American Board of Toxicology.

Rebecca Miksad, M.D., M.P.H., is engaged in health services and outcomes research as an attending gastrointestinal oncologist at Beth Israel Deaconess Medical Center (BIDMC) and as a senior scientist at the Institute for Technology Assessment at Massachusetts General Hospital. Her research has been supported by a Young Investigator Award from the American Society of Clinical Oncology. Her current and past research support includes the Timely Special Opportunity Award from the Dana-Farber Cancer Institute, the NIH Loan Repayment Program, and the Clinical Research Feasibility Fund Award from BIDMC. Recent awards include the Lee Lusted Prize for outstanding research from the Society for Medical Decision Making and the Clinical Research Award from the Eastern Cooperative Oncology Group.

Dr. Miksad received her B.A. in economics from Harvard University, an M.D. with honors in research from Cornell University, and an M.P.H. from Harvard University. She completed her internal medicine residency at New York-Presbyterian Hospital and her Hematology/Oncology fellowship at BIDMC. She completed the NCI-funded post-doctoral fellowship in the Dana-Farber/Harvard Cancer Center Program in Cancer Outcomes Research Training (PCORT). Dr. Miksad's research goals are to improve oncology treatment decision making through better characterization of cancer patient outcomes, improved accuracy of clinical endpoints, assessment of the economic implications of cancer therapy and application of decision analysis tools.

Arthur Schatzkin, M.D., M.P.H., Dr.P.H., received his B.A. from Yale University, his M.D. from the State University of New York Downstate College of Medicine, and an M.P.H. and Dr.P.H. from Columbia University School of Public Health. He completed residency training in Internal Medicine at Montefiore Hospital in the Bronx, NY. and Preventive Medicine at Mount Sinai Medical Center in Manhattan.. In 1984 he joined the National Cancer Institute, where he is currently chief of the Nutritional Epidemiology Branch, Division of Cancer Epidemiology and Genetics. Dr. Schatzkin's research in recent years has focused on prospective cohort studies of diet and cancer, with an emphasis on improving exposure assessment methods for investigating long-standing, but as yet unresolved, hypotheses (e.g., dietary fat versus breast cancer, fiber and fruits and vegetables versus colorectal cancer). He is a Principal Investigator for the NIH–AARP Diet and Health Study, a prospective cohort study of

diet and cancer among more than a half million U.S. men and women; the Polyp Prevention Trial, an intervention study of the effect of a low-fat, high-fiber, fruit- and vegetable-enriched diet on colorectal adenoma recurrence, and the Observing Protein and Energy Nutrition (OPEN) study, a biomarker-based investigation of the measurement error structure of dietary assessment instruments. He is currently exploring Internet- and metabolomics-based approaches to assessing nutrition–cancer relations in large prospective studies.

Robert Temple, M.D., is director of the Office of Medical Policy of the Food and Drug Administration's (FDA's) Center for Drug Evaluation and Research (CDER) and is also acting director of the Office of Drug Evaluation I (ODE-I). He has served in this capacity since the office's establishment in 1995. Dr. Temple received his M.D. from the New York University School of Medicine in 1967. In 1972 he joined CDER as a review medical officer in the Division of Metabolic and Endocrine Drug Products. He later moved into the position of director of the Division of Cardio-Renal Drug Products. In his current position, Dr. Temple oversees ODE-1, which is responsible for the regulation of cardio-renal, neuropharmacologic, and psychopharmacologic drug products. He also oversees The Office of Medical Policy, which is responsible for regulation of promotion through the Division of Drug Marketing, Advertising, and Communication and for assessing the quality of clinical trials. Dr. Temple has a long-standing interest in the design and conduct of clinical trials and has written extensively on this subject, especially on choice of control group in clinical trials, evaluation of active control trials, trials to evaluate dose–response, and trials using "enrichment" designs.

Aliza Thompson, M.D., M.S., is a medical officer in the Division of Cardiovascular and Renal Products of the FDA. She received her M.D. from Johns Hopkins University and completed her Internal Medicine and Nephrology training at Columbia University/New York-Presbyterian Hospital. She holds an M.S. in Biostatistics/Patient Oriented Research Track from the Columbia University Mailman School of Public Health.

Marc K. Walton, M.D., Ph.D., is currently associate director in the Office of Translational Science at the FDA's CDER. Dr. Walton received his graduate degrees from the University of Chicago. Later, he completed a medical internship at Rush University/ Presbyterian Medical Center in Chicago, followed by a neurology residency at University of Rochester. Following residency he moved to the National Institute of Neurological Disorders and Stroke, NIH, as a Senior Staff Fellow engaging in research on the development of neurotransmitter responses in the embryonic spi-

nal cord. In 1993, Dr. Walton joined the Center for Biologics Evaluation and Research (CBER) at the FDA as a medical officer. His work initially focused on clinical trials of investigational biological products (proteins, monoclonal antibodies, cellular therapies, and gene transfer therapies) as potential neurotherapeutic agents. He added the clinical areas of pulmonary, cardiovascular, endocrine, and hematologic disorders to his oversight when appointed to branch chief. Dr. Walton became the division director during the transfer of the biological protein product regulatory oversight from CBER to CDER, which broadened his areas of clinical oversight to all non-oncology uses of biological proteins. A subsequent move to the Office of Policy in the Office of the Commissioner gave him involvement in a broader range of agency-wide issues. He has now moved to the Office of Translational Science in CDER, where he is involved in fostering science and policies to support new approaches to therapeutic development.