Perspectives on Biomarker and Surrogate Endpoint Evaluation: Discussion Forum Summary

ISBN
978-0-309-16324-8

140 pages
6 x 9
PAPERBACK (2011)

Alison Mack, Erin Balogh, and Christine Micheel, Rapporteurs; Committee on Qualification of Biomarkers and Surrogate Endpoints in Chronic Disease; Institute of Medicine

PERSPECTIVES ON BIOMARKER AND SURROGATE ENDPOINT EVALUATION

Discussion Forum Summary

Alison Mack, Erin Balogh, and Christine M. Micheel, *Rapporteurs*

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
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.

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.

International Standard Book Number-13: 978-0-309-16324-8
International Standard Book Number-10: 0-309-16324-2

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). 2011. *Perspectives on biomarker and surrogate endpoint evaluation: Discussion forum summary.* 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.

COMMITTEE ON QUALIFICATION OF BIOMARKERS AND SURROGATE ENDPOINTS IN CHRONIC DISEASE

JOHN R. BALL (*Chair*), Senior Advisor, 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 DE GRUTTOLA, Professor and Chair, Biostatistics, Harvard School of Public Health

DAVID DEMETS, Professor, 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 HARLAN, JR., Consultant

ALLAN JAFFE, Professor of Medicine, Mayo Clinic

RONALD 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 NESS, Dean, University of Texas Health Science Center, School of Public Health

JENNIFER VAN EYK, Professor, Departments of Medicine and Biological Chemistry, Johns Hopkins University

JOHN A. WAGNER, Vice President, Clinical Pharmacology, Merck & Co., Inc.

Consultant

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

Study Staff
CHRISTINE M. MICHEEL, Study Director
SHARYL NASS, Senior Program Officer
ERIN BALOGH, Associate Program Officer
CAIRA WOODS, Christine Mirzayan Science and Technology Policy Graduate Fellow (January to April 2010), Research Associate (April to July 2010)
ASHLEY McWILLIAMS, Senior Program Assistant
PATRICK BURKE, Financial Associate
ROGER HERDMAN, Director, Board on Health Care Services
LINDA 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 process. We wish to thank the following individuals for their review of this report:

Douglas Balentine, Unilever
Thomas Fleming, University of Washington
Philip Greenland, Northwestern University
James Mayne, Pfizer, Inc.
Rebecca Miksad, Harvard Medical School
Jack Zakowski, Beckman Coulter, Inc.

Although the reviewers listed above have provided many constructive comments and suggestions, they did not see the final draft of the report before its release. The review of this report was overseen by **Sharon B. Murphy,** scholar-in-residence, Institute of Medicine. Appointed by the Institute of Medicine, she was 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 authors and the institution.

Contents

1

Introduction

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. (IOM, 2010)

Due to the absence of an agreed-upon process for biomarker evaluation, the Food and Drug Administration (FDA) requested that the Institute of Medicine (IOM) recommend a framework for the evaluation of biomarkers in the chronic disease setting and make ancillary recommendations for its application. In a report published in May 2010, the IOM Committee on Qualification of Biomarkers and Surrogate Endpoints in Chronic Disease (henceforth, "the committee") recommended such a framework, comprising three critical components: analytical validity, evidentiary qualification, and utilization analysis (IOM, 2010). This framework is intended to bring consistency and transparency to the previously nonuniform process of biomarker evaluation. The full summary of the committee's report is included in Appendix B.

On June 21 and 22, 2010, the IOM convened a 2-day discussion forum in Washington, DC, in order to provide an opportunity for stakeholders to learn about, react to, and discuss the report, *Evaluation of Biomarkers and Surrogate Endpoints in Chronic Disease* (IOM, 2010; see Appendix A). The discussion forum was attended by representatives of several FDA centers, including the Center for Food Safety and Applied Nutrition (CFSAN), the Center for Drug Evaluation and Research (CDER), the Center for Biologics Evaluation and Research (CBER), and the Center for Devices and Radiological Health (CDRH); by representatives of the National Institutes of Health (NIH), including the National Heart, Lung, and Blood Institute (NHLBI) and the Office of Dietary Supplements; by representatives of industry and industrial professional societies, including food, dietary

supplements, pharmaceuticals, medical diagnostics, and devices; and by representatives of public health, consumer, and nutrition consulting organizations, as well as by members of the committee and IOM staff.

Presentations reviewed the committee's work process, recommendations, and provided perspectives on the report from the point of view of participants from the FDA, the NIH, and from a diverse group of industry stakeholders; all such sessions were followed by panel discussions. Many presenters emphasized that the views they expressed were theirs and not necessarily those of their organizations or institutions. Thomas Fleming, professor of biostatistics and statistics at the University of Washington, gave a keynote presentation on the critical issues in the validation of surrogate endpoints, a specific use of a biomarker.

This document recounts the discussion forum proceedings, focusing in turn on each represented sector. A summary of Dr. Fleming's presentation then sets the committee's report within the context of biomarker utilization. Lastly, this summary examines the main themes raised by stakeholders, and the challenges and opportunities presented to stakeholders by the report's recommendations.

2

Committee Findings and Recommendations

CHARGE TO THE COMMITTEE

As noted in the report (IOM, 2010), the Center for Food Safety and Applied Nutrition (CFSAN), in conjunction with the Food and Drug Administration's (FDA's) Center for Drug Evaluation and Research (CDER), approached the Institute of Medicine (IOM) in 2008 for advice on the topic of biomarker and surrogate endpoint evaluation in chronic disease. These FDA centers expressed concern regarding the limited number of surrogate endpoints available, the high cost of evaluating possible surrogate endpoints, and the absence of an agreed-upon, systematic, and transparent process for biomarker evaluation. They also wished to learn whether principles of biomarker qualification or evaluation learned in the drug development setting could be applied in other FDA-regulated product categories, such as foods. CFSAN thus requested that the IOM charge an expert committee with the following tasks:

> 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)[1] such as

[1] The terms *coronary heart disease* and *cardiovascular disease* are often used interchangeably. In this report, the use of either term reflects the speaker's choice.

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]

On the basis of this statement of task, the committee undertook its work to address these charges as ensured by the IOM external review process, said committee chair John R. Ball, senior advisor at the American Society for Clinical Pathology.

Committee Process and External Review Process

The IOM convened the committee, which comprised experts from a variety of related fields and was supported by a highly capable staff, said Dr. Ball. The committee met in person four times and had several teleconferences over the course of a year in order to fully develop their charge, set a plan of work, gather relevant evidence, and develop its findings and recommendations. The committee benefited from presentations by outside experts in a workshop format and from comments offered by interested parties. As part of its charge, the committee reviewed alternate biomarker evaluation models; this review can be found in Chapter 2 of the committee's report (IOM, 2010).

The committee's report underwent a rigorous external review, which helped focus and clarify their findings and recommendations. Fourteen reviewers participated in this process, and two individuals appointed by the National Research Council and the IOM oversaw the review, according to Dr. Ball.

Early in committee deliberations, Dr. Ball recalled that the committee recognized that "biomarkers are really useful when used carefully." Biomarkers have served a variety of diverse uses, he noted, which include

- Discovery and development of medical therapies and products,
- Comparative effectiveness research,
- Formation of clinical practice guidelines,

[2] The terminology in the statement of task differs in a few ways from the terminology adopted by the committee, which replaced *qualification* with *evaluation* in many instances, and *risk biomarker* with *biomarker*.

- Basic biomedical research,
- Clinical practice,
- Public health practice, and
- Understanding healthy nutrition and lifestyle choices.

However, Dr. Ball continued, the committee also quickly recognized that a biomarker's usefulness is strongly dependent upon context. "No single biomarker is good for everything," he said, which became a defining principle that informed the committee's recommended framework for biomarker evaluation, discussed below. This framework and additional supporting recommendations were introduced in the workshop's first session, following the definition of a series of terms relevant to biomarker evaluation and application. This session also reviewed examples of biomarker case studies to which the committee applied their evaluation framework.

DEFINITIONS

As noted in their report (IOM, 2010), "The committee observed a great deal of inconsistent and imprecise definition and use of terms relevant to biomarkers and biomarker evaluation." Having determined that consistent, precise definition and use of terms is critical for biomarker evaluation, the committee strove to be both consistent with the spirit of previous efforts in this vein and also to clarify several potentially confusing definitions. The results of this process, which provided a foundation for the committee's task, were presented at the workshop's outset by committee member John A. Wagner, vice president of clinical pharmacology at Merck & Co., Inc.

Biomarker

Dr. Wagner noted that the Biomarkers Definitions Working Group (BDWG), convened by the National Institutes of Health (NIH), produced the definition of a biomarker that is widely used today (Biomarkers Definitions Working Group, 2001). The IOM committee's version of this definition is: "a characteristic [for example, cholesterol level] that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a[n] … intervention." The committee used this definition in its report, and further defined "objectively" to mean "reliably and accurately" in this context.

Risk Biomarker and Risk Factor

A related term, *risk biomarker,* was used by the FDA's CFSAN in their request to the IOM to charge an expert committee to "generate recommendations on the qualification process for biomarkers, with a focus on risk biomarkers and surrogate endpoints in chronic disease" (IOM, 2010). The committee defined a risk biomarker to be "a biomarker that indicates a risk factor for a disease," which includes genetic biomarkers, Dr. Wagner said. He noted that this definition contrasts with that used previously by CFSAN, which characterizes risk biomarkers as "biological indicators that signal a changed physiological state that is associated with the risk of a disease" (CFSAN, 2009), and therefore does not include genetic biomarkers.

Dr. Wagner said that the committee resolved to use the term *biomarker* instead of *risk biomarker* in order to clearly delineate between biomarkers and *risk factors,* which are defined in the report as "variables that predict outcomes and are composed of biomarkers and social and environmental factors" (IOM, 2010). As noted in the report, the value of a risk factor depends on its ability to predict an event.

Surrogate Endpoint and Clinical Endpoint

The widely accepted definition of *surrogate endpoint* was proposed by the BDWG in 2001, Dr. Wagner said. According to this definition, a surrogate endpoint is "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" (Biomarkers Definitions Working Group, 2001). For example, blood pressure has served as a surrogate endpoint for morbidity and mortality due to cardiovascular disease (CVD) in trials of several classes of antihypertensive drugs, Dr. Wagner said.

A surrogate endpoint represents a special use of a biomarker, in which the biomarker substitutes for a clinical endpoint. Closely following the BDWG definition, the committee defined *clinical endpoint* as "a characteristic or variable that reflects how a patient [or consumer] feels, functions, or survives" (Biomarkers Definitions Working Group, 2001). Death is one example of a clinical endpoint.

Dr. Ball and Dr. Wagner noted important examples of successes and failures of biomarkers that have been used as surrogate endpoints. Two key successes are blood pressure as a surrogate endpoint for CVD clinical endpoints, and HIV-1 RNA as an indicator of complete viral suppression for HIV interventions. By contrast, arrhythmia suppression proved a fail-

ure as a surrogate endpoint for interventions meant to reduce cardiac sudden death; similarly, low-density lipoprotein cholesterol (LDL-C) reduction through hormone replacement therapy failed to provide a surrogate endpoint for CVD clinical endpoints.

RECOMMENDATIONS

The committee concluded that focusing solely on biomarker qualification—the process of determining whether a biomarker of interest is associated with a specific clinical endpoint—would not sufficiently address the committee's charge, said Dr. Wagner. The committee saw their primary task as identifying a process for biomarker evaluation, which they fulfilled in large part by recommending a three-part framework comprising

1. Analytical validation, which asks the question, is the biomarker able to be accurately measured?
2. Qualification, which asks the question, is the biomarker associated with the clinical endpoint of concern? and
3. Utilization, which asks the question, what is the specific context of the proposed use?

Dr. Wagner said that the first two pieces of the framework, analytical validation and qualification, are already commonly accepted; on the other hand, he noted, "the addition of utilization is a little bit more controversial, but nonetheless extremely useful." He added that the committee envisioned all three elements of biomarker evaluation as interdependent components of an iterative cycle, as illustrated in Figure 2-1.

The committee's recommendations are shown in Box 2-1. In accordance with their charge, the committee applied the biomarker evaluation framework to a series of case studies and made additional recommendations for implementing the framework, for supporting evidence-based decision making, and for promoting public health. Presentations by three committee members reviewed the background for and rationale behind each recommendation, and also presented in detail the committee's case study of low-density lipoprotein cholesterol (LDL-C) and high-density lipoprotein cholesterol (HDL-C) as biomarkers for cardiovascular risk, one of several such studies that appear in their report. These presentations were followed by a discussion session open to all workshop participants.

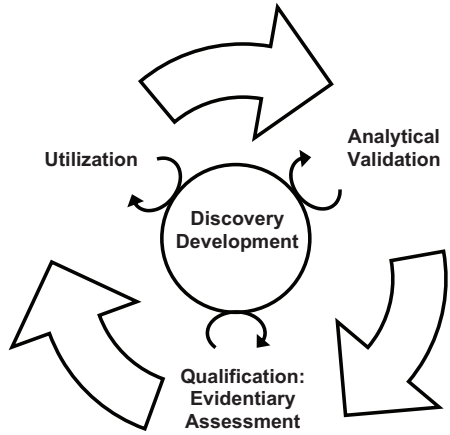

FIGURE 2-1 The steps of the evaluation framework are interdependent. While a validated test is required before qualification and utilization can be completed, biomarker uses inform test development, and the evidence suggests possible biomarker uses. In addition, the circle in the center signifies ongoing processes that should continually inform each step in the biomarker evaluation process.
SOURCE: IOM, 2010.

Recommendations 1 and 2: Biomarker Evaluation

Dr. Wagner discussed each component of the biomarker evaluation framework described in Recommendation 1 (see Box 2-1): analytical validation, qualification, and utilization. Analytical validation comprises analyses of the available evidence regarding the analytic performance of a particular assay or biomarker; this, he said, is a necessary first step to determining how a biomarker is performing. He defined *qualification*, a term which has in some cases been confused with *validation*, as the assessment of the available evidence that links a biomarker with a biological process, such as a disease state, a clinical outcome, or an intervention. Utilization, a concept subsumed under *qualification* in some previous biomarker evaluation frameworks, is a distinct process in the committee's evaluation framework because it represents a subjective and contextual analysis—as contrasted with the objective analyses in the analytical validation and qualification steps—specific to the use of a given biomarker or surrogate endpoint, Dr. Wagner said. He added that the committee conceived of the three steps of biomarker evaluation as components of an interactive and iterative cycle, as illustrated in Figure 2-1.

An analysis of the many sources of variability in biomarker measurements—which include biological, sample collection, and analytical factors—led the committee to conclude that "biomarker tests need to

BOX 2-1
Recommendations

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.

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.

SOURCE: IOM, 2010.

be reliable, reproducible across multiple laboratories and clinical settings, and maintain adequate sensitivity and specificity before data based on them can be used in subsequent evaluation steps," Dr. Wagner said. This finding provided a particular focus and a foundation for the committee's work.

Tumor Size and Analytical Validation

Per their charge, the committee used examples whenever possible to illustrate their findings. Dr. Wagner used the case study of tumor size—described in detail in the committee's report on pages 135–142 (IOM, 2010)—to highlight the role of analytic validity in biomarker use. Measurements of tumor size are crucial in determining the efficacy of cancer therapeutics and in daily clinical practice, he said. However, tumor size can be defined by many different technologies (and different techniques within those technologies). Tumor size can be gauged by measuring tumor diameter, volume, or mass; these in turn are measured by a variety of platforms and techniques that include magnetic resonance imaging (MRI), computed tomography (CT), and positron emission tomography (PET) imaging, Dr. Wagner said. Further, different contrast agents and different protocols may be used, all of which affect measurement precision. Thus, he concluded, "analytical validation of tumor size is complicated by multiple imaging platforms and other assay performance issues."

CRP and Qualification

Biomarker qualification requires analysis of the nature and the strength of evidence for the relationship between a given biomarker and a disease-associated biologic pathway, and of evidence that interventions targeting the biomarker affect the clinical endpoints of interest, Dr. Wagner said. To illustrate this process, the committee examined C-reactive protein (CRP), a biomarker that has been shown in observational studies to serve as an independent predictor of future vascular events, including myocardial infarction (MI), ischemic stroke, peripheral vascular disease, and vascular death. The committee's application of the biomarker evaluation framework to CRP found evidence for its prognostic value, but insufficient support for its use as a surrogate endpoint. As Dr. Wagner noted, CRP is a useful biomarker of CVD, but its utility as a surrogate endpoint has not been established (see also pages 142–153 of IOM, 2010).

Troponin and Utilization

The utilization component of the biomarker evaluation framework is used to determine whether the analytical validation and qualification conducted on a given biomarker provide sufficient support for a specific proposed use, Dr. Wagner explained. He emphasized that strong evidence and a compelling context are needed for the use of a biomarker as a surrogate endpoint—that is, as a substitute for the clinical endpoint.

This requirement is illustrated by the case of troponin, a biomarker used ubiquitously in acute settings to diagnose MI. Troponin can be elevated due to a variety of chronic heart conditions, inflammatory conditions, side effects from drugs, or organ failures, Dr. Wagner said. While there is evidence that prevention of MI reduces death rates, none supports the proposition that using an intervention specifically to decrease troponin levels improves mortality risk (IOM, 2010). Thus, he concluded, evidence is lacking to support the use of troponin as a surrogate endpoint for interventions in these situations (see also pages 153–159 in IOM, 2010).

Dr. Wagner acknowledged that decisions made regarding biomarker utilization are necessarily subjective, in contrast to qualification, which he characterized as a "data-gathering and evidentiary step." This difference was a key rationale for separating qualification and utilization, he said (see also Recommendation 2b. in Box 2-1). However, he added, it is also the case that the three distinct elements of the biomarker framework are interrelated and may be pursued concurrently; moreover, the results of one evaluation step (for example, utilization) may reveal the need for revisions or additional work in one or both of the other steps (for example, analytical validation and/or qualification).

Beta-Carotene and the Biomarker Evaluation Process

The committee's second recommendation (see Box 2-1) focuses on the evaluation by the FDA of biomarkers and biomarker tests with regulatory impact. The three parts of this recommendation specify that (1) such evaluations should be conducted by expert panels convened by the FDA; (2) analytical validation and qualification of biomarkers should be evaluated separately from utilization; and (3) biomarkers should be continually evaluated on a case-by-case basis.

Dr. Wagner offered the example of blood levels of beta-carotene, proposed as a biomarker for risk of CVD and cancer, to illustrate the need for an expert panel to periodically evaluate evidence associated with a particular biomarker. Years of epidemiological studies showed that diets rich in fruits and vegetables were associated with lower incidence of CVD and cancer, leading many to believe that beta-carotene was responsible

for the lower risk, he said. "However, definitive clinical trials showed that this hypothesis was incorrect and that supplementation with beta-carotene did not lower risk for cancer or cardiovascular disease," and in some cases, it increased risk (see also pages 168–175 of IOM, 2010). These circumstances suggest the need for periodic evaluation of the evidence for the analytical validation, qualification, and utilization of biomarkers with regulatory impact, he concluded.

LDL and HDL Case Study

Committee member Ronald Krauss, director of atherosclerosis research and senior scientist at the Children's Hospital Oakland Research Institute, presented an in-depth analysis of the committee's fifth case study, LDL- and HDL-cholesterol as biomarkers for cardiovascular risk, to illustrate the application of the biomarker evaluation framework (see also pages 159–168 in IOM, 2010).

The context for the LDL- and HDL-cholesterol case study, illustrated in Figure 2-2, is the relationship between both LDL-C and HDL-C and CVD risk, as established by the Framingham Heart Study (Castelli, 1988). LDL-C and HDL-C are, respectively, components of LDL and HDL particles. As Dr. Krauss noted, LDL comprises multiple subclasses of particles with differing composition. Research in many populations has affirmed that LDL-C is directly related to cardiovascular risk, independent of HDL-C, and that HDL-C is inversely related to cardiovascular risk. These findings led to the widespread use of LDL-C and HDL-C as biomarkers for evaluating the efficacy of pharmacologic and nutritional interventions for CVD. The committee's case study evaluated each of these biomarkers in various contexts of use.

"Clearly, LDL stands as one of the major FDA-qualified surrogate endpoints for cardiovascular disease," Dr. Krauss stated. LDL-C is often viewed as a benchmark biomarker, particularly for nutritional claims, he said, despite the fact that LDL-C is but one component of LDL particles, which vary in composition. "The evidence supporting LDL as a biomarker rests almost entirely on the measurement of LDL-C," he concluded. He also noted that in some populations, LDL particle measurements provide better estimates of CVD risk than LDL-C.

Along with these issues, the committee considered the disease context for evaluating LDL-C or LDL particles as biomarkers, Dr. Krauss said. CVD, he noted, is a very complex disease that is increasingly recognized as a spectrum of pathologic and pathophysiologic effects, only one of which is primarily related to progression of the cholesterol content of plaques as a function of LDL-C in the blood. Atherosclerosis, he added, is "often indolent, progressive over time, and then is complicated by a

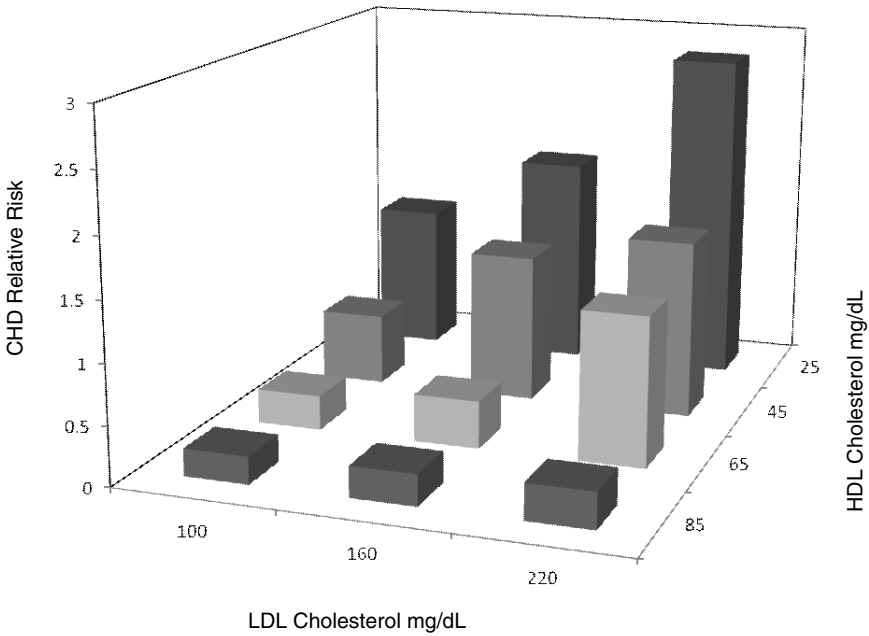

FIGURE 2-2 Relative risk of coronary heart disease after 4 years compared to several LDL-C to HDL-C ratios. Men aged 50–70 years in the Framingham Heart Study.
NOTE: HDL-C = high-density lipoprotein cholesterol; LDL-C = low-density lipoprotein cholesterol; mg/dL = milligrams per deciliter; RR = relative risk; Y = years.
SOURCE: Castelli, 1988. Adapted, with permission, from the Pulsus Group, 2010. Adapted from the Canadian Journal of Cardiology 1988;4(SA):5–10.

number of additional factors that can convert a cholesterol-rich plaque to a more malignant form that destabilizes and is involved with both inflammation and thrombosis; immune changes can occur that could be critical." Thus, he concluded, "it is rather simplistic to consider either LDL or HDL, or even the two of them together, as sufficient to explain these complex mechanisms."

The predictive value of LDL-C for CVD events varies considerably as a function of health status, Dr. Krauss said. As shown in Figure 2-3, risk for cardiovascular events associated with high LDL-C in patients without diabetes and CVD was found to be significantly lower than LDL-C-associated event risk in patients with both conditions (Robinson and Stone, 2006). Likewise, he noted, when the effects of LDL-lowering interventions (for example, statins) were compared in a meta-analysis

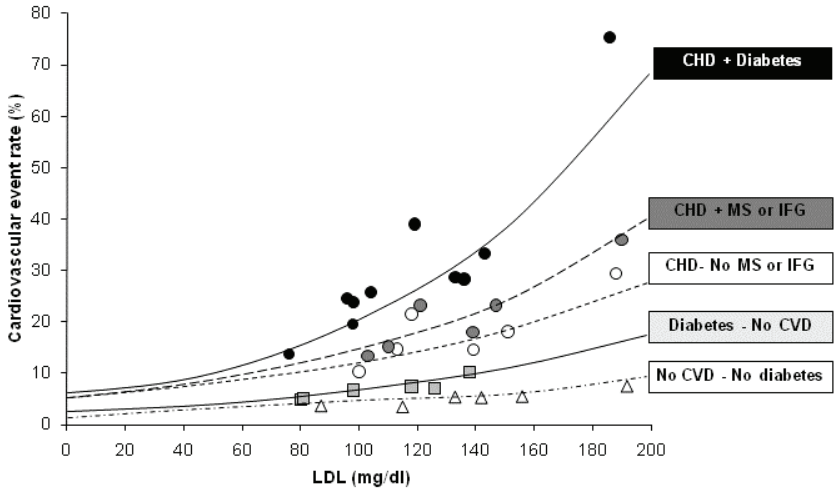

FIGURE 2-3 CVD risk varies over wide range of LDL-C.
NOTE: CHD = coronary heart disease; CVD = cardiovascular disease; IFG = impaired fasting glucose; MS = metabolic syndrome.
SOURCE: Robinson and Stone, 2006. Reprinted, with permission from Elsevier, 2010. Copyright 2006 by Elsevier.

(Baigent et al., 2005), some studies were found to deviate (but not significantly) from the generally linear relationship between reduction in LDL and major coronary events. "This was not unexpected, given the multiplicity of mechanisms involved, both in the disease process and in the mechanism of drug action," Dr. Krauss observed.

"One of the classic examples of failure of LDL as a surrogate endpoint—and it actually applies to HDL as well—[is] the lesson from hormone replacement therapy," he said. Postmenopausal hormone replacement therapy (HRT) was first thought to protect women from CVD based on both observational epidemiological data and the apparent beneficial effects of estrogen on both LDL and HDL and other biomarkers for CVD, he explained. "Fortunately," he continued, "a number of people in the field held out for clinical trials, and some major clinical trials demonstrated that HRT had no benefit on CVD incidence in healthy women, and actually increased mortality in the first year of treatment in women with preexisting cardiovascular disease and increased the risk for thromboembolic events."

Dr. Krauss summarized the conclusions of the committee's case study of LDL-C as a biomarker of CVD risk as follows:

- The strength of LDL-C as a surrogate endpoint is not absolute due to the heterogeneity of CVD processes, the heterogeneity of LDL-lowering drug effects, and the heterogeneity of LDL particles themselves.
- Age, gender, and genetic factors have been shown to complicate the already complex dynamics of the LDL–CVD relationship; as a result, lowering LDL-C can never be considered a "perfect" indicator across all population groups.
- Nonetheless, there is high probability that lowering LDL-C by certain interventions (for example statins) decreases risk of CVD, and LDL-C, although not perfect, is one of the best biomarkers for CVD.

Dr. Krauss then turned briefly to HDL-C and its potential as a biomarker for CVD risk based on the inverse relationship demonstrated in the Framingham Heart Study (Castelli, 1988; Gordon et al., 1977). "There is strong epidemiologic evidence for a relationship of HDL to cardiovascular risk, and there's also quite a bit of pathophysiologic evidence indicating that certain therapeutic maneuvers, most of them based on genetic manipulations in animal models, can raise HDL to reduce atherosclerosis progression or disease risk in these models," he stated. In humans, however, he noted that HDL "is even more heterogeneous than LDL and includes multiple subpopulations of particles with differing functional properties and disparate effects on atherogenic mechanisms."

"There are multiple ways of raising HDL cholesterol and HDL particle concentrations, lifestyle being one of them," Dr. Krauss continued. "But a number of drugs can raise HDL through different mechanisms, and understanding those drug effects simply by using HDL as an index of benefit does not address a fundamental question: is the HDL being raised by a process that would confer the expected benefit [reducing CVD risk]?" Thus, he concluded, "there is as yet no conclusive evidence in humans for an independent benefit of HDL increase on CVD outcomes, in large measure because trials aimed at raising HDL almost always have concordant effects on other cardiovascular risk markers that cannot easily be teased apart."

As an example of these findings, Dr. Krauss presented the results of a large trial of the first HDL-C raising drug, a cholesteryl ester transfer protein inhibitor called torcetrapib (Barter et al., 2007). This trial demonstrated that torcetrapib had the capacity to raise HDL-C by over 70 percent from baseline over the course of the first year of treatment, and also to reduce LDL-C levels, he said. Nevertheless, he continued, despite the favorable apparent effect on these two CVD biomarkers, "the overall result of the trial showed not just the absence of a benefit, but actually a higher overall

mortality in individuals treated with this drug in combination with statin compared with statin alone." This result "clearly represents a complex outcome that is not just a failure of the surrogate endpoint," Dr. Krauss said. "We now know that there are off-target effects of this drug on cardiovascular disease itself, not manifested by any biomarker yet that we clinically use, that could have accounted for the adverse effect."

Dr. Krauss then discussed the committee's case study of LDL and HDL as CVD biomarkers in the context of a well-established paradigm for evaluating relationships between disease and intervention, surrogate endpoints, and clinical outcome, devised by committee member David DeMets, professor of biostatistics and medical informatics at the University of Wisconsin, and Thomas Fleming, who gave an overview presentation later in the discussion forum (see Chapter 7 in this volume) (Fleming and DeMets, 1996). As noted by Fleming and DeMets (1996), "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."

This can happen, for example, if the surrogate endpoint is not in the causal pathway of disease. "We don't have a real example of that here," Dr. Krauss said, "because for the most part, the LDL and HDL measurements are in the causal pathway [for CVD] to some degree or another." However, he added, one could argue that depending on the way by which LDL and HDL are measured (for example, as LDL-C and HDL-C), these biomarkers may not in fact reflect clinical outcomes.

Another scenario for biomarker failure under the paradigm occurs when an intervention (for example, statin) affects a biomarker (for example, LDL-C) favorably, contributing to improved clinical outcome, Dr. Krauss said. However, he observed, in this case the result "is not a perfect biomarker for the outcome of a trial, because the disease process itself has an extraordinary effect on the clinical outcome that's independent of the intervention."

In a third case of biomarker failure, the intervention affects a clinical outcome independently of a surrogate endpoint, according to Dr. Krauss. He offered the example of a diet intervention such as the Mediterranean diet in the Lyon Diet Heart Study, in which interventions improved clinical outcome without having any obvious benefit on LDL, HDL, or other surrogate endpoint (de Lorgeril et al., 1999).

Finally, a biomarker can fail when an intervention has multiple effects on the outcome itself, on the surrogate endpoint, and on the relationship of the disease to the outcome, Dr. Krauss stated; this is the case in the two previous examples involving the effects of torcetrapib and HRT on CVD. In these cases, "there may have been a benefit through the pathway of lowering LDL and raising HDL, but it was obscured by other effects of the drug and … of the disease process," he said.

"In the end, biological complexity above and beyond simple measurements that we're considering here lead to many opportunities for error," Dr. Krauss concluded. As shown in Figure 2-4, interventions may influence disease outcomes through multiple pathways; as he noted, "this is certainly the case for LDL and HDL … which comprise particles arising through different pathways that may have different pathological effects, leading to different outcomes." Thus, he concluded, the main lesson the committee learned from the LDL and HDL case study is that "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."

Dr. Ball added that this and the other case studies, among other evidence considered by the committee, led them to conclude that biomarkers "are good in many circumstances, but that it depends largely on the context of use." He followed this with an anecdote from his own experience, which appears in the preface to the report. "Several years ago," he said, "I had three episodes of atrial fibrillation, and after the third one, I realized that all three of those episodes had been associated with the drinking of two glasses of red wine." Upon recognizing this correlation, he stopped drinking red wine, and thereafter has not had another episode of atrial fibrillation.

However, when he told his mother about his conclusions, she replied, "but I thought red wine was good for your heart." As Dr. Ball learned, however, "it depends." Red wine, he said, may be "good for the plumbing, the coronary arteries, apparently, but it was not so good for my pacemaker, the electrical system." Similarly, he observed, sometimes an intervention affects biomarkers, sometimes it affects the disease itself, and sometimes it doesn't have an effect—or it has a negative effect—on clinical outcomes. These circumstances, he said, led the committee to devise the three-part biomarker evaluation framework comprising analytic validation, qualification, and utilization, depending on context of use.

Discussion and Clarification of the Biomarker Evaluation Framework

In the discussion period that followed the committee presentations, several committee members sought to affirm that the focus of their work, per their charge, was to develop an evaluation framework for biomarkers, and not for FDA-regulated products or interventions (for example, drugs, biologics, medical devices, foods, or nutritional supplements) that might be identified or characterized through the use of biomarkers. Presenter Andrew Shao, senior vice president of scientific and regulatory affairs at the Council for Responsible Nutrition, raised this issue when he asked whether the committee had considered the case of biomarkers

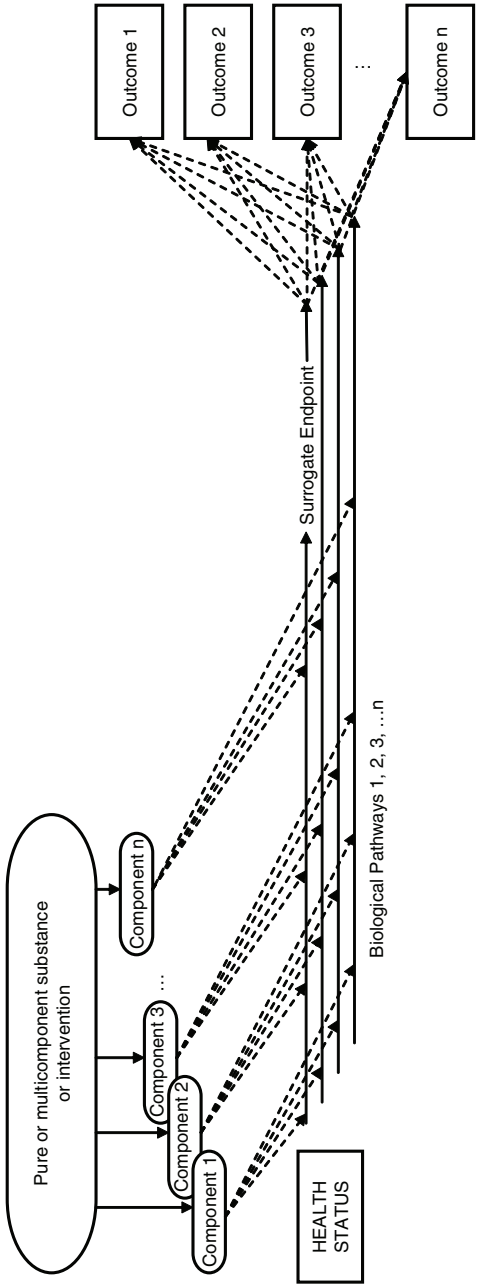

FIGURE 2-4 Multiple ingredients, multiple biological pathways, and multiple outcomes illustrate some of the complexities of the use of biomarkers and surrogate endpoints in chronic disease. Note that while the solid horizontal arrows indicate biological pathways, they do not necessarily indicate pathways 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.
SOURCE: IOM, 2010.

for nutrient exposure, which inform policies such as dietary guidelines, and if so, whether Recommendation 3 should be interpreted to mean that nutritional guidance should be developed according to the paradigm currently used for drugs.

"This committee was charged with evaluating the criteria for qualifying biomarkers themselves," Dr. Krauss responded. "That's a different question, I think, from the one you asked, because that information [biomarker qualification] can then be used by policy makers, dietary guidelines formulators, et cetera, in the context of the strength of the data, the strength of the basis for that qualification," he explained. The goal of biomarker evaluation is to provide firm evidence on which such policy decisions can be made, he added.

The committee's task was to identify criteria to be used for judging whether a given surrogate endpoint represents an appropriate way to monitor a specific disease or biological process, according to committee member Jennifer Van Eyk, professor in the departments of medicine and biological chemistry at Johns Hopkins University. The committee was not focused on the various entities that a biomarker or surrogate endpoint may be used to test, she said; instead, they considered the criteria by which surrogate endpoints should be judged and the challenges of developing surrogate endpoints that can be used throughout the regulatory process.

Some workshop participants sought further clarification of the distinction between utilization and qualification, as pertains to the recommended evaluation framework, and also between their use of the terms *biomarker* and *surrogate endpoint*. Presenter Marc Walton, medical officer at CDER, asked whether a biomarker (for example, HDL-C) determined to be inappropriate for use as a surrogate endpoint for a particular clinical endpoint (for example, CVD), can be said to have failed the qualification step as well as the utilization step.

Dr. Ball explained that the qualification step answers two questions: whether there is a relationship between the biomarker and the clinical endpoint, and whether the intervention of interest affects both the biomarker and the clinical endpoint in the same way. If that's the case, utilization answers whether a particular context of use for the biomarker—which may or may not be as a surrogate endpoint—makes sense. "It probably is not possible to qualify HDL as a biomarker for everything," Dr. Ball concluded. "The context of use matters."

Indeed, Dr. Krauss pointed out, HDL failed on the second part of the qualification step because in certain situations—some of which he described in his presentation—an improvement in HDL-C levels did not improve disease outcomes (for example, when the intervention of interest is HRT). Clearly, HDL-C is a biomarker for CVD in a general sense,

he said, but it's an imperfect biomarker in the sense that it doesn't represent a component of HDL most responsible for its relationship to CVD. "One of the holy grails in our field is to identify a functional test for HDL that could actually correlate fairly consistently with efficacy in drugs or diets in reducing risk [for CVD] through measurement of that functional property of HDL," he observed; such a test would replace HDL-C and, if it demonstrated a more direct connection with disease outcome, conceivably could achieve qualification and utilization as a biomarker.

"The analytical validation and qualification pieces are fit for the intended purposes," Dr. Wagner added. The analytical validation and qualification requirements for a particular biomarker used in research would not be the same if that biomarker was to be used for a regulatory purpose; for example, although there are qualification issues with tumor size as a biomarker in oncology, it would be appropriate for use as a biomarker for evaluating a candidate anticancer agent under certain circumstances, especially given the dearth of cancer biomarkers. The committee recognized that the decision to use a biomarker for a regulatory purpose is a subjective one, he reported, so they separated utilization from qualification.

Utilization is the most challenging component of the biomarker evaluation framework from a policy-making perspective, noted committee member Roberta Ness, dean of the University of Texas School of Public Health. How to analyze and characterize a biomarker through testing are well understood, she observed, but the committee also recognized that many biomarkers work perfectly well in particular contexts, while failing in other contexts. "The parameters around the 'it depends' are complicated … and some of them are perhaps even counterintuitive," she said. For instance, when assessing a biomarker in the context of a homogenous ill population for whom few medical interventions exist, Dr. Ness suggested that one might be much more willing to accept an imperfect biomarker as a surrogate endpoint, as a means to identifying better interventions, than in the context of a food consumed by the general population.

"My understanding of what the committee is saying is if one has a wealth of data … from clinical studies that [for example] establish that treatment effect on LDL is predicting treatment effect on death from MI, to what extent can we actually then implement that in the real world?" Dr. Fleming stated. "I would think the answer to that [question] would come in utilization," he continued, "where you would say, if you had other classes of agents that were similar to those that were the source of the data, that extrapolation would be reasonable. But if you had a new, novel intervention such as … torcetrapib, that could readily be influencing off-target effects, one would be much more reserved about that degree of generalization."

The same notion is true of blood pressure, Dr. Fleming added. As he noted in his presentation, considerable data show that effects on blood pressure predict effects on stroke for wide classes of agents, including low-dose diuretics, angiotensin-converting enzyme (ACE) inhibitors, beta blockers, calcium channel blockers, and angiotensin II receptor blockers (ARBs). In the case of a drug with a novel mechanism of action, however, "one would have more reservation about simply labeling [it] for stroke based on [its] effect on blood pressure." Thus, he concluded, in the utilization step, one takes "the totality of the data from the analytical validation and the critically important qualification steps into a real-world setting to [ask], to what extent can we generalize those known relationships to determine whether the biomarker could be used in the specific instance as a basis for judging clinical benefit?"

On a more practical level, presenter Guy Johnson, principal of Johnson Nutrition Solutions, LLC, asked how biomarker utilization would be executed by the FDA. Given the agency's long history of using advisory committees, Dr. Ball noted, it would be appropriate for the FDA to assemble groups of individuals with relevant expertise—including stakeholders with a range of perspectives—to address specific cases. These advisory committees would first assess the data on validation and qualification as they pertain to the biomarker under review, and then examine the particular use for which the biomarker is proposed.

Recommendations 3 and 4: Scientific Process Harmonization

Dr. Ness introduced the remaining recommendations in her presentation. While these recommendations, like the evaluation framework, apply to all products regulated by the FDA (i.e., drugs, medical devices, biologics, foods, and dietary supplements), Dr. Ness said she would focus her remarks on foods because, as she explained, this regulatory area was central to the committee's charge.

Dr. Ness summarized Recommendation 3 (see Box 2-1) with the statement, "all regulatory areas should be considered equal." She explained that the rationale for this recommendation came from the committee's recognition that a very large number of people in the population may be exposed to any given food, and thereby, any toxic effects that might result from its consumption. "I think we lull ourselves into the idea that foods cannot be toxic, but clearly there are good examples where that's in fact not the case," she said, invoking Dr. Ball's anecdotal experience with red wine, as well as medical allergies to foods, which have both broad and severe impact. In addition, she noted, currently, "there is no learned intermediary between a person and their consumption of food."

Recommendation 4 (see Box 2-1) sends the message that "we should

not be considering biomarkers as a single element within a complex foodstuff and, in fact, within a complex diet," Dr. Ness stated. Rather, she continued, "we need to really consider the whole food when talking about a particular relationship between a biomarker and an outcome." Returning to the scenario depicted in Figure 2-4, Dr. Ness pointed out that a food is an example of a "pure or multicomponent substance or intervention," illustrating the complex relationship that can exist between a food or individual components in food, biomarker measurements, and clinical outcomes. Therefore, she said, the committee concluded that to "minimize the uncertainty inherent when biomarkers are used to predict beneficial effects of a food substance, the substance's effect should be evaluated in its context of use—the whole food product and dietary patterns associated with use."

In the discussion period that followed this session, audience member Elaine Krul, molecular nutrition lead of Solae, LLC, observed that drug efficacy is often considered in the context of diet; for example, she said, "if you're taking a statin, be careful when you take your grapefruit juice; when you're taking warfarin, be very careful with your background diet and … vitamin K intake; if you're taking synthroid, you have to be careful not to take other foods that inhibit the absorption of the drug." However, she added, to extend such a scientifically rigorous analysis to foods would be difficult. Dr. Krauss replied that while it is important to examine the health effects of foods within the context that they are consumed, "the science is never going to be perfect." He acknowledged a fundamental lack of certainty based on the science at hand, but asserted that probabilities of such effects should be determined, considered, and communicated to both medical professionals and consumers in ways that clearly convey their inherent uncertainty.

Referring to a publication considered to be a classic in the biomarker literature, Dr. Ness noted that Prentice (1989) not only insisted that biomarkers fully reflect clinical outcomes, but also that they must capture the entirety of the effect of an intervention on an outcome (for more detail about the Prentice criteria see pages 27 and 56 of IOM, 2010). "Clearly, if one takes just one component out of a complex foodstuff, one could not capture with a biomarker that whole complexity with respect to outcome," she added. This viewpoint prompted considerable discussion. Presenter Stephen Williams, chief medical officer of SomaLogic, Inc., noted that analytical validation and qualification of biomarkers can always be improved, and the consequences of error are uncertain, and asked, "how good is good enough?" Dr. Wagner replied that the committee recommended the continuous and iterative evaluation of biomarkers.

Dr. DeMets also commented that the Prentice biomarker criteria are applied with consideration to a biomarker's context of use, and that a

biomarker used as a surrogate endpoint must capture the effect of the intervention of interest, whether it is a medicine, a device, a food, or a supplement. "If you really are serious about evaluating the effect of any new intervention, from devices and drugs to foods, it really is an essential step," he said. The failure of biomarkers such as LDL and beta-carotene to predict clinical endpoints resulted from incomplete knowledge of the biology of complex diseases such as cancer and CVD, he said. "The committee struggled with the issue of biomarkers versus clinical endpoints and the fact that they are all measurements, and they are all subject to measurement issues," Dr. Wagner added.

Dr. Ness noted that the FDA has, in the case of several foods and supplements shown in Table 2-1, accepted health claims that do not reflect the entirety of a foodstuff. On the other hand, she continued, "just recently the FDA has decided that there are claims that are not acceptable to them." For example, she said, the agency recently informed General Mills that their claim that eating Cheerios can lower cholesterol by 4 percent in 6 weeks exceeded the regulatory framework for food and would trigger a recategorization of this product as a drug. Dr. Ness reported that in March 2010, the FDA notified 17 food manufacturers regarding food labels that violated the Food, Drug, and Cosmetics Act by making unauthorized health claims, nutrient claims, and through their use of the word *healthy*.

Presenter Douglas Balentine, director of nutrition sciences at Unilever, Foods, noted that at present, very few foods make FDA-approved, ingredient-based health claims, and that most foods containing ingredients listed in Table 2-1 are marketed on the basis of general dietary guidance (for example, "Diets rich in fruits and vegetables may reduce the risk of some types of cancer").

Ancillary Recommendations 5 and 6: Improving Evidence-Based Regulation

Dr. Ness introduced the committee's ancillary recommendations intended to support the implementation of the biomarker evaluation framework across all areas regulated by the FDA (see Recommendations 5 and 6 in Box 2-1). According to Dr. Ness, the first ancillary recommendation strengthens recommendations made by previous panels, including the IOM committee that authored the report, *The Future of Drug Safety* (IOM, 2007b). These groups advised that the FDA be granted additional authority related to postmarket surveillance in areas beyond its existing focus for such studies: drugs and biologics.

TABLE 2-1 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.

Postmarket Studies of Products Approved on the Basis of Surrogate Endpoints

"The current situation is that postmarket studies can be required [by the FDA] for drugs and biologics under certain circumstances, and in particular when there [is] an accelerated approval based upon a surrogate endpoint," Dr. Ness stated. For example, for a drug approved for adults and subsequently used in pediatric population, or for a drug approved on the basis of animal studies because clinical trials in humans were unethical. For devices, she said, "postmarket surveillance is challenging due to a lack of reliable monitoring of adverse events and lack of publicly accessible information." Surveillance of the effects of claims on foods is limited, and the process for removing harmful claims can be very slow, she added.

Dr. Ness observed that the FDA has recently strengthened its authority with respect to postmarket surveillance for drugs, and expressed hope that such surveillance could be extended to other regulatory areas, particularly medical devices and foods. The committee determined that surveillance of biomarker-based food health claims is needed, she said, because it is important to understand the impact of the actual foodstuffs on consumers. "When surrogate endpoints are used in product or claim approvals, data collection is needed to link the product or claim to health outcomes experienced by patients and consumers, whether for drugs, biologics, devices, foods, or supplements," she said. The committee's Rec-

ommendation 5a acknowledges that in order to achieve this goal, it will be necessary to adapt the FDA's existing regulatory frameworks, which currently differ among product types, Dr. Ness explained.

In subsequent discussion, Dr. Williams expressed confusion as to the current state of the FDA's authority to demand postmarket surveillance of the products it regulates. Noting that the committee's report states that the enactment of the Food and Drug Administration Amendments Act (FDAAA) in 2007 granted the agency "new authorities to require postmarket studies" that formerly would have been voluntary (IOM, 2007a, p. 207), he asked what gap in current legislation Recommendation 5a was intended to fill. Responding to Dr. Williams' question, study director Christine M. Micheel, of the IOM, said that with little evidence to date as to whether the FDAAA effectively addressed the committee's concerns, they wanted to emphasize the need for legislation—whether existing or strengthened—to require postmarket studies of products approved on the basis of surrogate endpoints. Dr. Ball added that the committee considered evidence from the FDA and others that some postmarket studies requested by that agency had not been undertaken by industry; whether these failures resulted from the FDA's lack of authority or resources remained unclear, he said.

Responding to Dr. Balentine's request for a description of an acceptable process for fully qualifying a food health claim, Dr. Ball observed that while the committee felt that the same scientific rigor ought to be brought to bear on such claims as on drugs or biologics approved on the basis of surrogate endpoints, they also recognized the need for practicality. "When pharmaceuticals go on the market, there's a patent, and there's a huge financial incentive for information to be developed and brought before the FDA," he said. Because foods lack patents and the incentives they provide, there are practical difficulties in funding the gathering of information on foods. This gap, he said, is addressed in Recommendation 6 (discussed below), which states that the Department of Health and Human Services (HHS) should develop mechanisms for collecting and sharing data on biomarkers through public–private collaborations. "I think we do recognize a difference in the capabilities of the food industry and the capabilities of the pharmaceutical industry [to collect data]," Dr. Ball concluded, "but at the same time … when a claim is made on the basis of surrogate endpoints, there needs to be follow-up, because there can be unintended consequences of the claim."

Given the committee's position regarding the need for studies of food health claims based on surrogate endpoints, audience member Chor San Khoo, vice president of global nutrition and health at Campbell Soup Company, Inc., wondered whether the committee would recommend similar evaluation of diets. For example, "The DASH [Dietary Approaches to

Stop Hypertension] diet clearly is a very effective diet in reducing blood pressure … [and] it's a composition of many foods … [so is gathering data related to use of that diet] also necessary?" Dr. Ness and Dr. Krauss responded that such a recommendation would exceed the committee's charge, which pertained to products regulated by the FDA on the basis of biomarkers. The real question, Dr. Krauss added, is whether existing biomarkers are sufficient to assess the health effects of complex diets. "For any individual [biomarker], such as blood pressure, I think that the intervention can be a complex diet, with the understanding that there's going to be other effects of that diet," he said, "but blood pressure as a biomarker for one of those effects is legitimate."

Consumer Understanding of Claims on Foods and Supplements

The second part of Recommendation 5 (see Box 2-1) derives from the committee's recognition of a general lack of consumer understanding of biomarker-based claims on foods and supplements, according to Dr. Ness. For these products, "consumers must evaluate information without the advice of a learned intermediary in most circumstances, [so] consumer understanding of health information based on biomarkers is critically important," she said. Dr. Ball added, "if a statement is made with regard to a biomarker, but the public has difficulty in understanding that, that's an issue that we believe that FDA should examine and that industry should examine, as well." The committee thus concluded that the FDA "needs the authority to request or require studies of consumer understanding, to continue conducting its own research in this area, and the authority to remove or require changes to claims or other information that are not understood by consumers," Dr. Ness said.

The committee spent considerable time talking about the complex issue of consumer understanding of health claims, Dr. Ness said. Numeracy is a serious concern, as is the tendency of consumers to misunderstand food labels—particularly those that appear on the front of a package, which may cause them to ignore the required nutritional information provided on the side or back. This tendency has been reflected in research that further suggests that consumers have difficulty distinguishing among the various evidentiary levels of health claims on food labels and understanding the qualifying language that these claims frequently contain. The committee therefore concluded that the FDA needs the authority to request and require studies of consumer understanding and to make changes or remove claims that are misunderstood by consumers.

Recommendation 6 (see Box 2-1) supports the committee's previous recommendations by encouraging the collection and sharing of biomarker-related data across the HHS, and in particular, the creation of

coordinated data infrastructure and surveillance systems, Dr. Ness said. To this end, she noted that the FDA has been building a program known as the Sentinel Initiative,[3] a national electronic system for the postmarket safety tracking of drugs, biologics, medical devices, and eventually, all FDA-regulated products. She observed that ClinicalTrials.gov, which provides publicly accessible information about clinical trials, could provide "important guidance for biomarker data collection efforts" as well. Dr. Ness also recognized the following current and ongoing initiatives that may be brought to bear on Recommendation 6:

- Biomarkers Consortium[4]—a public–private partnership to identify, develop, and qualify biomarkers;
- Critical Path Institute[5]—an independent, publicly funded organization that supports the FDA's Critical Path Initiative by working with industry, regulatory agencies, and academia on a range of projects, including the development of biomarkers;
- CEO Roundtable on Cancer Life Science Consortium[6]—a precompetitive collaborative effort that is working with the National Cancer Institute (NCI) to select promising cancer biomarkers for development; and
- Oncology Biomarker Qualification Initiative[7]—a collaboration between the NCI, FDA, and the Centers for Medicare & Medicaid Services (CMS).

While noting that each of these initiatives occupies a specific niche, she said that all of them could be more cohesively and uniformly leveraged, "bringing together data sources and partners to provide the best research infrastructure to further evaluate the relationship between interventions, biomarkers, and outcomes across regulatory areas."

[3] See http://www.fda.gov/Safety/FDAsSentinelInitiative/default.htm.
[4] See http://www.biomarkersconsortium.org/.
[5] See http://www.c-path.org/.
[6] See http://ceo-lsc.org/.
[7] See http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/2006/ucm108597.htm.

3

FDA Perspectives

The Food and Drug Administration (FDA) must balance competing needs in its approach to evaluating biomarkers, and it must do so with limited resources (IOM, 2010). On one hand, biomarkers are viewed as a route to reducing the cost and time required to develop effective new drugs, devices, and biologics to address chronic diseases; on the other hand, patients and consumers must be protected from risks associated with biomarker use. The apparent efficiency/safety dichotomy also applies to the use of biomarkers in health claims for foods, which the FDA regulates under a different framework than is applied to drugs, devices, and biologics.

These complexities were apparent in workshop presentations and discussion reflecting the FDA's perspective on the report. Speakers representing three FDA centers involved in using biomarkers in regulatory decision making—Center for Food Safety and Applied Nutrition (CFSAN), Center for Drug Evaluation and Research (CDER), and Center for Devices and Radiological Health (CDRH)—were asked by the committee to describe the process their centers currently employ to evaluate biomarkers, to compare that process with the committee's recommended evaluation framework, and to consider how the FDA might make use of their report. To the last question, Marc Walton replied that at the time of the workshop, the report had only recently been published, and would require detailed reading by several people at each FDA center involved with biomarker evaluation before the agency could respond to it. "We're clearly very interested in the report and are going to be thinking about it

very carefully," he said. "It's clearly a very extensive, detailed, effortful report … so there will be a lot that we want to think about."

Discussion following the presentations focused on two main topics. The first was the definition of the term *risk biomarker*, as used in the two presentations from representatives of CFSAN and employed in its charge to the committee. As previously noted, the committee did not use this term, due to its potential for confusion with the precisely defined term *risk factor*. Elizabeth Yetley, a consultant to the National Institutes of Health (NIH) Office of Dietary Supplements and to the committee, asked whether CFSAN considered a risk biomarker to be a type of surrogate disease biomarker. Speaker Paula Trumbo, supervisory biologist at CFSAN, replied, "risk biomarkers can be surrogate endpoints, but not all risk biomarkers are surrogate endpoints."

This exchange prompted Thomas Fleming to ask whether CFSAN's goal was to be able to determine whether a given biomarker is a reliable way to assess the level of risk a patient has for a certain event, or whether they seek biomarkers that can represent reliably whether a treatment will alter a patient's risk for such an event, and can therefore serve as a surrogate endpoint. "The former [case] simply requires a biomarker to be a correlate," he said, but the latter requires the effect on the biomarker to reliably predict the full effect on the true clinical endpoint. Dr. Fleming further explored this distinction in his presentation, which is summarized in Chapter 7 in this volume.

A second topic of discussion focused on whether a surrogate endpoint must have biological plausibility: that is, that the biological connection between the biomarker and the process it represents is known, if not fully elucidated. Stephen Williams noted that, in an article he and John A. Wagner had written recently reviewing the history of surrogate endpoints, they had argued against requiring biological plausibility for surrogate endpoints (Lathia et al., 2009). They reasoned that highly plausible biomarkers had failed as surrogate endpoints (for example, ventricular arrhythmia, as previously discussed by Ronald Krauss), and also that this criterion would exclude some types of biomarkers (for example, microRNA) for which plausibility might be difficult or impossible to determine.

However, Dr. Williams added, an editorial in the same issue of the journal in which their article was published disputed their conclusion (Gobburu, 2009). Therefore, he asked the FDA representatives how they viewed the role of biological plausibility as a criterion for selecting surrogate endpoints. The committee's report did not specifically recommend that surrogate endpoints have biological plausibility, Dr. Williams noted.

"From my point of view, plausibility is always helpful," Dr. Walton stated. "For one, it provides enough enthusiasm to expend the significant

amount of energy and effort it takes to prove [a surrogate] endpoint," he said. "If it seems to be just magic, it can be hard to sustain the effort over time." He also said that thinking about the apparent mechanism underlying a biological process helps researchers plan experiments or studies to test the endpoint in question.

From a theoretical perspective, Dr. Walton agreed that a biomarker with an unknown link to a biological process, but with robust supporting data, could have utility as a surrogate endpoint; however, he noted that it might not be clear whether this endpoint reflected a late, critical stage in the disease process, or an early step in the response to an intervention. Moreover, he said, if a biomarker's biological significance is known, that information might suggest further applications for the biomarker, leading to the testing of additional hypotheses and the possible expansion of its use. Without a mechanistic understanding of the biomarker's biological role, its use would be restricted to a specific, data-constrained context.

"It really is important to do the best we can to objectively understand the data and formulate hypotheses," Dr. Fleming said. According to Dr. Fleming, biological plausibility is particularly critical to formulating a hypothesis regarding a putative surrogate endpoint; thereafter, prospective trials are needed to validate the effect of an intervention on the biomarker as compared to a clinical endpoint. "Now, in that validation trial, I will also have had prespecified biological mechanisms that will be key supportive endpoints, and those can, in fact, be more readily interpreted as confirmatory," he explained, "yet I'll explore the data even in those trials … but those additional exploratory analyses of biological mechanism have to be viewed with great caution."

Committee member Victor De Gruttola, professor and chair of the Department of Biostatistics at Harvard School of Public Health, noted that the association between a given surrogate and a clinical endpoint may be confounded. Given that possibility, unless the underlying biological mechanism is understood, "there isn't really a statistical way to be sure that confounding has been appropriately dealt with," he said. "Surrogacy analyses, especially the ones that attempt to ascertain whether the Prentice condition holds or not, will always be subject to confounding." Therefore, he said that clinical trials must be conducted on surrogate endpoints in order to assess causality and determine the impact of a randomized treatment on an endpoint.

PRESENTATION BY PAULA TRUMBO, CFSAN

Dr. Trumbo reviewed the current process for biomarker evaluation at CFSAN, which occurs in the course of reviewing petitions for authorized and qualified health claims for foods. Scientific evidence offered in

support of such claims often includes academic studies employing "risk biomarkers [that] are not validated surrogate endpoints for chronic disease risk," she reported.

A health claim is a causal relationship between a substance, which can be a food or a food component, and a disease or a health-related condition, Dr. Trumbo said. Such claims apply to healthy people in the U.S. population who are free of (but possibly at high risk for) the disease that is subject to the health claim. For example, she said, a person who has elevated low-density lipoprotein cholesterol (LDL-C) levels could be included in an evaluation of a heart disease health claim. Health claims can also be crafted to address a particular subpopulation such as women or the elderly, she added.

Health claims are not about treating, preventing, curing, or mitigating symptoms of a disease, Dr. Trumbo said, which differentiates them from drug claims. "Certainly, there are drugs that are available for reducing modifiable risk factors, i.e., surrogate endpoints," she said. "But when we evaluate the evidence for a health claim, we're either looking at a surrogate endpoint or a clinical disease endpoint … we won't be looking at evidence that is about mitigation of a disease … [such as] progression of an existing cancer or joint pain from osteoarthritis." The statutory laws governing the regulation of foods and drugs are very different, said Dr. Trumbo. Provided they are free of pathogens, foods are generally recognized as safe (GRAS), and CFSAN focuses its premarket safety evaluations on determining whether ingredients that are added to foods or to dietary supplements are safe.

CFSAN evaluates health claims according to an evidence-based review system that has been in place for about 5 years, according to Dr. Trumbo. This system establishes scientifically rigorous criteria for two categories of health claims—authorized health claims and qualified health claims. Dr. Trumbo noted that what differentiates these categories is the strength and quantity of the evidence supporting them. A qualified health claim must be presented to consumers using qualifying language because the evidence behind the claim is not as strong as that supporting an authorized health claim, which does not require qualifying language.

Evidence for health claims includes intervention studies, clinical studies, and observational studies. Dr. Trumbo noted that CFSAN considers intervention studies to be their gold standard, particularly those that measure a clinical endpoint. More often, however, CFSAN must evaluate evidence for health claims based on observational studies involving surrogate endpoints. "For instance, one of our qualified health claims is on monounsaturated fats and olive oil, and another one … [is for] walnuts and coronary heart disease," both of which utilized LDL-C data to support the claims.

In evaluating the strength of evidence presented to support health

claims, CFSAN considers the type, size, and quality of studies conducted and their relevance to the general or target population of the claim, said Dr. Trumbo. In addition to surrogate endpoint biomarkers, CFSAN also encounters biomarkers of intake (for example, beta-carotene and other carotenoids) as part of health claims. In this case, she said, "if needed, we will do side reviews to evaluate whether there is good evidence to suggest that an intake biomarker is a good reflection of intake."

Dr. Trumbo presented the following list of disease endpoints CFSAN has been petitioned to review, coupled with associated biomarkers:

- Coronary heart disease (CHD)—Total/LDL cholesterol, blood pressure
- Colon/rectal cancer—polyps
- Diabetes—Blood sugar levels, insulin resistance
- Osteoporosis—Bone mineral density
- Dementia—Mild cognitive impairment

None of these endpoints has been determined to be a surrogate endpoint by CFSAN, she said. "We don't have the expertise to make that decision, so … we rely on the various institutes within the National Institutes of Health, as well as our colleagues at CDER" to evaluate surrogate endpoints and keep CFSAN informed of new developments in this area. However, CFSAN came to wonder whether they could apply similar criteria to the qualification of biomarkers as they do to the health claim review process. This question prompted CFSAN to approach the Institute of Medicine (IOM), and is the basis for their request for a biomarker evaluation framework, said Dr. Trumbo.

PRESENTATION BY KATHLEEN ELLWOOD, CFSAN

Kathleen Ellwood, director of nutrition programs staff in the Office of Nutrition, Labeling, and Dietary Supplements in CFSAN, further outlined the challenges that led to CFSAN's request for the IOM study. She noted that there are few known "validated modifiable risk biomarkers for chronic disease risk," and "not all risk biomarkers are surrogate endpoints." In particular, she said, many chronic diseases lack surrogate endpoints.

Filling this gap will not be easy if it requires reliance on costly long-term clinical trials using clinical outcomes, said Dr. Ellwood. Therefore, the use of "biomarkers and biomarkers of disease risk" is important to CFSAN, especially in the chronic disease setting. However, there is "an absence of an agreed-upon, systematic, and transparent process for qualifying surrogate endpoints," and a need for research to identify potential surrogate endpoints and the scientific evidence needed to support their

use. "We need to identify these information gaps, including the type of additional studies, [and] the level of evidence that's needed to qualify a risk biomarker," she said.

Dr. Ellwood noted that a 2007 IOM report funded by CDER recommended that government agencies and other stakeholders develop "a transparent process to create well-defined consensus standards and guidelines for biomarker development, validation, qualification, and use" (IOM, 2007a). As a result, Dr. Ellwood and her colleagues at CFSAN worked extensively with their counterparts at CDER to develop the request for this study, which was funded through the FDA's Critical Path Initiative.[1] To achieve their goal of developing "a framework on evidentiary standards for chronic disease risk biomarkers," CFSAN asked the IOM to recommend "a process, focusing on risk biomarkers and surrogate endpoints in chronic disease, and … asked [the IOM] to look at existing prototypes."

CFSAN requested that the IOM committee demonstrate application of the biomarker evaluation framework through case studies, as reflected in the committee charge. Dr. Ellwood noted that they wanted to focus the scope of the case studies to biomarkers and surrogate endpoints for CHD, a relatively well-researched area. For example, she noted that "LDL cholesterol is considered a surrogate endpoint, but HDL-C [high-density lipoprotein cholesterol] has not quite gotten to that level. So what would it take for HDL, then, to be considered a surrogate endpoint for a chronic disease?" CFSAN's objective in requesting that the IOM recommend a framework and demonstrate its use through case studies was "to assist FDA in developing a framework for the evidentiary standards for the qualification of biomarkers as potential surrogate endpoints of chronic disease risk."

PRESENTATION BY MARC WALTON, CDER

CDER's interest in biomarkers rests primarily in their ability to improve drug development and to improve CDER review decision making. "If all works well, then we'll wind up with more drugs that get successfully developed, and those that do get developed may be better optimized than without the biomarkers," said Dr. Walton. Within this general purpose, he identified the following ways in which biomarkers could help optimize drug development:

[1] The Critical Path Initiative is a strategy, launched by the FDA in 2004, to drive innovation in the scientific processes through which FDA-regulated products are developed, evaluated, and manufactured (http://www.fda.gov/ScienceResearch/SpecialTopics/CriticalPathInitiative/ucm076689.htm).

- Selection of candidates for development
- Study design during development
- Marketing application review
- Labeling information
- Enabling otherwise infeasible studies to be conducted (for example, surrogate endpoints)

As a result of the importance of biomarkers within CDER, Dr. Walton noted that the center is working actively with various groups to aid them as they try to develop the biomarkers and the data to support their use.

To date, biomarkers have become accepted for use within CDER through a two-stage process, according to Dr. Walton. Initially, a biomarker is proposed for use within a specific drug application, when a sponsor asks to use a particular biomarker in the drug approval process. The sponsor will discuss with CDER the data supporting their request, and then CDER makes a case-specific decision as to whether the biomarker's use is appropriate. "As these cases occur time and again in slightly different situations, over an extended period of time" CDER gains experience with the biomarkers, including a broader concept of the breadth of their use that is reliable and informative," Dr. Walton said. Based on that deeper experience, he said, CDER "will come to accept biomarkers in a broader sense."

CDER continues to refine and expand biomarker qualification along two main tracks, Dr. Walton continued. In some cases, biomarker tests will need to be codeveloped to evaluate a biomarker along with the drug whose use will be tied to that biomarker test result, he said; guidance for that process is being developed in collaboration with CDRH. In addition, Dr. Walton noted that CDER has established the Biomarker Qualification Process, which has been in development and slowly maturing over the past couple of years. CDER is currently developing a guidance for this program.

"Within CDER, when we talk about biomarker qualification, we are talking about a conclusion that within some very carefully and specifically and clearly stated context of use, this biomarker has been demonstrated to be able to support appropriate decision making," Dr. Walton stated. Once a biomarker is qualified, CDER is careful that it is being used within this specific context, and that new evidence has not since come to light that contradicts its qualification. As long as that is the case, drug developers can employ the biomarker without extensive or detailed submission of primary data to support its qualification, said Dr. Walton. "Drug developers will be able to rely upon that [biomarker] qualification, and the CDER reviewers will also be able to rely upon it."

When CDER qualifies a biomarker, "what we qualify is the substance,

or the analyte, that is … being measured," Dr. Walton said. "It is not the method that measures it, nor is it the device that measures the biomarker." Rather, CDER expects that a robust biomarker would have several different assay methods capable of measuring it appropriately, along with information on the performance characteristics of each of these assay methods. In its capacity to review devices for clinical testing, CDRH would clear or approve some of these assay methods, he said, but such clearance is not equivalent to CDER qualification of the biomarker being tested.

Dr. Walton characterized "context of use" as a "shorthand term for a comprehensive statement of the manner and purpose of use of the biomarker." He noted that context of use might, for example, be restricted to certain disorders or drug classes, or for nonclinical biomarkers to certain animal species. Context of use is also likely to specify how clinical samples must be obtained, he continued, although sample handling methods are considered to be part of the assay (and as such, guided by CDRH).

"Very importantly, how the results get interpreted and applied must be clearly stated within the context of use," Dr. Walton stated. For CDER, "a context of use defines the boundaries of the known reliability of the biomarker," he said. "It's not going to at all define the boundaries of nonreliability, because there are going to be a great many potential uses that have not [yet] been evaluated." Those uses are candidates for case-by-case considerations, he added, and if successful, for expanding the qualified context of use for that biomarker.

After reviewing the multistep biomarker qualification process currently practiced at CDER (see Box 3-1), Dr. Walton noted that the guidance being developed for this process will be incorporated into a more general guidance for "drug development tools (DDTs)." These include patient-reported outcome tools as well as biomarkers, he said, because their development and evaluation processes are quite similar. The guidance will describe the process of working with CDER to develop DDTs, explain CDER's managed process for consistent scientific evaluation of these tools, and establish the legal framework for CDER to post its qualification decisions themselves as guidance, he said. "It is important to us that this [guidance] becomes available for public access," he said. In addition, as with all guidances that FDA publishes, this allows for public comment regarding specific biomarkers and CDER's evaluation process.

Regarding the committee's recommended biomarker evaluation framework, Dr. Walton noted that, in some elements, the conceptual organization of the process is somewhat different from that used at CDER. He emphasized the importance of context of use, which he said must be clearly identified and specified in order for CDER to determine whether or not a biomarker is determined to be appropriate for use (*qualified* in CDER's terminology). He added that the analytical method should be

BOX 3-1
CDER Process of Biomarker Qualification

- Submission of a short letter requesting initiation of formal interactions with Biomarker Qualification Program
  - Subject to CDER acceptance of request
- Advice and consultation stage begins
- Submission of a briefing document
  - Identification of biomarker
  - Intended context of use
  - Existing state of knowledge
  - Identified knowledge gaps
  - Proposed approach to fill the gaps
- Meeting with Biomarker Qualification Review Team
  - Discussion, and advice on plans
- Repeated interactions as needed by sponsor to complete work of developing comprehensive evidence on biomarker for specific intended use
- Meeting with Biomarker Qualification Review team when data complete
  - Comprehensive evidence overview assessment
  - Agreement that all reasonably apparent knowledge gaps eliminated
  - Evidence supportive of qualification for the intended context of use
- Qualification review stage begins
- Submission of full, comprehensive, detailed evidence package with primary data
- Biomarker Qualification Review team evaluates data, writes review(s)
  - Makes recommendations on qualification decision
  - Repeated interactions as needed by biomarker qualification reviewers
  - Public advisory committee or other public consultation if warranted
- Letter of qualification sent to sponsor when CDER-wide concurrence
- Qualification letter and reviews to be posted on CDER website
  - After CDER Process Guidance published

SOURCE: Walton presentation (June 21, 2010).

reliable, reproducible, and adequately sensitive, and fit for the intended purpose. Dr. Walton noted that "for some purposes, a method [of measuring a biomarker] may have adequate sensitivity, and for other purposes, it may not be adequately sensitive." Therefore, "the validity of the method and its performance characteristics have to be thought about with the context of use in mind."

PRESENTATION BY FEDERICO GOODSAID, CDER

Federico Goodsaid, chemist at CDER, began by echoing Dr. Walton's emphasis on context, as well as remarks by Dr. Williams, who argued

that absent a context of use, it is difficult (if not impossible) to evaluate a biomarker.

"We have had a process for qualifying biomarkers over the past 100 years," Dr. Goodsaid said. "You wait long enough until you think you have all the references in the world needed to convince someone that a biomarker is valid. And when people become a little tired of arguing about it, they say, 'okay, it's qualified,'" Dr. Goodsaid said. "I think we can do better."

The real problem with qualification is defining the evidence required to qualify a biomarker for a particular use, Dr. Goodsaid said. For example, if one examined some of the biomarkers presented as case studies in the report without prior knowledge of their shortcomings, "what would we ask of a qualification process to conclude that they are qualified?" He said this is the overall question the FDA would like to be able to address, and it led him to further inquire, "how do you convince people that a biomarker is qualified for a specific purpose? What level of evidence is needed? If it is going to be for safety, what are the levels of evidence? If it is going to be for efficacy, what are the levels of evidence? Does it matter?"

There have not been many discussions regarding biomarker qualification for nonclinical uses, Dr. Goodsaid noted. He mentioned one exception: an account of a multiyear effort that resulted in the qualification of seven renal safety biomarkers by the FDA and the European Medicines Agency (EMEA) (Dieterle et al., 2010), based on the submission of drug toxicity studies and analyses of biomarker performance from the Predictive Safety Testing Consortium (PSTC). He described this achievement as a first step toward the goal of getting new biomarkers as efficiently as possible with comprehensive data supporting their claims.

In closing, Dr. Goodsaid made the point that, unless biomarker qualification offers immediate financial gain for companies, motivation will be lacking for the development and qualification of new biomarkers. "We need to think about the fact that many qualification efforts will be closely linked to consortia that will be pulling their resources together to be able to try to qualify biomarkers," he argued, and offered as examples the PSTC and an effort by the Chronic Obstructive Pulmonary Disease (COPD) Foundation. According to Dr. Goodsaid, the ultimate goal is to qualify biomarkers that are needed for a broad range of applications and to do so as quickly as possible, provided sufficient data are available to support specific contexts of use.

Dr. Goodsaid noted that a sensible approach to making better and wider use of biomarkers would begin by qualifying various biomarkers for limited contexts of use, followed by attempts to expand their uses based on additional evidence. However, he added, the context of use must be known in order to know what data would be needed to support that application of the biomarker.

PRESENTATION BY ROBERT BECKER, CDRH

Robert Becker, medical officer at CDRH, began his presentation by commenting on the IOM report: "It is notable, from my perspective, to see the extensive and nuanced assessment of biomarkers that has been provided in the report. We recognize that this is a challenging scientific area, and there is very obviously admirable product with respect to the effort and the report itself that has been generated."

He then provided an overview of device regulation at CDRH, which include some biomarker tests. CDRH regulates devices for treatment, diagnosis, or any other application with a medical context. Like drugs, regulation focuses on safety, which he defined as the reasonable assurance that the probable benefits of a device will outweigh its risks, and effectiveness, which he defined as the reasonable assurance that use of the device will prove clinically significant.

Medical devices are classified by CDRH according to the level of risk associated with their use, said Dr. Becker. In the case of biomarkers, the risk is the potential harm to patients if the biomarker does not accurately reflect the clinical endpoint. Uses deemed high risk are often associated with cancer diagnosis, he explained, or with large population exposure, as occurs for some screening procedures (for example, mammography). In these cases, CDRH requires a premarket application, which establishes the safety and effectiveness of the individual device for its intended use, he said. Devices classified as moderate risk include products used to determine cancer prognosis and monitor diagnosed cancer in patients; these require a premarket notification by the sponsor. By contrast, low-risk devices such as tongue depressors and toothbrushes are usually exempt from premarket submissions.

In response to an audience question, Dr. Becker noted that biomarker tests can also be approved through another pathway defined by the Clinical Laboratory Improvement Amendments (CLIA).[2] "There are, of course, two paths by which tests used for medical care have made their way to market over the past 3 decades or so," Dr. Becker said. While the FDA approval process focuses on the test itself, the CLIA regulations focus on the laboratory in which the tests are conducted, and are aimed at ensuring the quality of its procedures, staffing, qualifications, and management.

[2] The Centers for Medicare & Medicaid Services (CMS) regulates all laboratory testing (except research) performed on humans in the United States through the Clinical Laboratory Improvement Amendments (CLIA). In total, CLIA covers approximately 200,000 laboratory entities. The Division of Laboratory Services, within the Survey and Certification Group, under the Center for Medicaid and State Operations (CMSO) has the responsibility for implementing the CLIA Program. The objective of the CLIA program is to ensure quality laboratory testing. See http://www.cms.gov/clia/.

Over the past several years, the FDA has expressed an increased interest in regulating aspects of laboratory development tests, he noted, adding that this is a topic of active discussion.

A growing area of involvement by CDRH concerns the regulation of biomarkers intended to inform the course of medical treatment, according to Dr. Becker. A well-known example of such "companion diagnostics" is a test used to determine whether a patient with breast cancer could benefit from receiving the drug trastuzumab, an antagonist against a specific receptor (human epidermal growth factor receptor [HER2/neu]) that is overexpressed by certain breast cancers. "This is, I think, a prime illustration of the kind of circumstances that will tend to underlie personalized medicine as it evolves," said Dr. Becker. Moreover, he added, when personalized therapeutic decisions are driven by single test results, as is becoming increasingly common, the risks associated with that test are increased due to the heightened consequences of its failure. "You are not talking about a circumstance where the test result is simply integrated with a lot of other information," he said. "You are now talking about a circumstance where the test result itself can drive a clinical decision."

For this reason, Dr. Becker emphasized that companion diagnostics should be reviewed and approved by the FDA before the test goes to market, because "the risks for such tests equal the risks of the therapeutic products." He added that under such circumstances, CDRH aims to optimize the benefit-to-risk performance associated with using the test and treatment together.

Dr. Becker commended the report's treatment of analytical validation, and said that it reflected many of the considerations CDRH has in evaluating biomarker tests. He provided a list of these considerations, which he called "not complete by any means, but … typical of the kinds of features or performance characteristics" that enable CDRH to assess how well a test measures a biomarker of interest:

- Precision (repeatability, reproducibility);
- Accuracy;
- Sensitivity, limit of detection;
- Specificity (interference, cross-reactivity);
- Sample type/matrix;
- Sample preparation/conditions;
- Performance around the cutoff; and
- Potential for carryover, cross-hybridization.

The steps CDRH takes to conduct analytical validation of a biomarker-based device vary according to the technologies involved, said Dr. Becker. For example, different statistical paradigms may be used to evaluate

qualitative and quantitative tests, or the evaluation paradigm may reflect the setting in which the biomarker test is used. Typically, these tests are kits or devices intended for laboratory use, but CDRH has also regulated tests conducted by a single laboratory.

A test's intended use strongly influences its evaluation by CDRH, according to Dr. Becker. Context of use determines which features are critical to device design; these features may include the analyte itself, how it is measured, the population in which the test will be used, and the rationale for using the test. Ideally, this information is explicitly stated in the course of defining its intended use in the approval or clearance process, but he noted that applicants to CDRH have "a long history of … varied adherence to clarity" in the information they provide to support such devices. This shortcoming, he added, may reflect the fact that the science of device regulation is relatively new, compared with drug regulation, and also that some devices have more uses than most individual drugs. Thus, the center finds that their regulatory paradigms are continually evolving as their experience with the uses and risks of devices grows, he added.

Dr. Becker presented what he called an "unofficial" list of elements that inform device regulation by CDRH:

- The description of the product,
- Its intended use (clinical context, purpose, target population, means of usage),
- Instructions for its use,
- Performance claims and limitations (for example, is it prognostic versus predictive? How well does it fulfill its intended purpose?),
- How it will be manufactured and marketed, and
- How potential problems with the device may be detected and resolved.

Dr. Becker pointed out that these devices may operate in a range of medical contexts that include risk estimation, screening, diagnosis, residual or refractory disease, recurrent disease, monitoring, prognosis, and prediction of therapeutic response. He emphasized that each application is unique in the studies or trials required to demonstrate the analytical and clinical validity of the test.

It is often desirable to have more than one device on the market to meet a single intended use, said Dr. Becker. This is particularly the case for biomarker-based tests, he added, because they can employ a variety of measurement methods. When a "follow-on" test is developed for a biomarker for which a qualified test already exists, it raises several issues, Dr. Becker said. "Even in something as simple as an immunohistochemistry

test for HER2, you can have antibodies that might be associated perhaps with different epitopes of the antigen that might convey with that somewhat different performance characteristics associated with the test," he said. "You can find, for example, that the very same analyte … might show up in a different clinical context" (for example, two different tumor types that vary in their expression of the analyte).

Using HER2 as an example, Dr. Becker offered further explanation of considerations regarding "follow-on" tests based on biomarkers. HER2 is expressed at a variety of levels that can be assessed from the perspective of either biochemistry or cellular physiology, with one method measuring the HER2 protein itself while the other method relies on reverse-transcriptase polymerase chain reaction (RT-PCR) to monitor expression of the HER2 gene. "These different approaches to measuring the biomarker bring with them different aspects of analytical performance," he said, to the extent that "you wonder at what point they actually diverge in terms of the[ir] biological implications." In addition, two tests might measure the same biomarker in different matrices (for example, within a tumor versus serum levels).

The science associated with biomarkers and their applications is iterative and continually evolving, and it is important to recognize that scientific understanding of a biomarker can be "outrun by technical progress," said Dr. Becker. For example, gene signatures represent a descriptive or empirical compilation of biomarkers that might be shown to correlate with clinical outcomes or features, but for which the underlying biological significance is unknown. He noted that an ongoing, resource-intensive process of biomarker qualification and utilization, as recommended by the committee, could lead to increased understanding of the biological role of each biomarker and the expansion and refinement of its uses.

4

National Institutes of Health Perspectives

As noted in the report (IOM, 2010), the National Institutes of Health (NIH) has played an instrumental role in advancing biomarker discovery, development, and qualification. The agency has initiated a range of efforts, including workshops and consortia, aimed at improving stakeholder collaboration and expanding public access to information on promising biomarkers. The report recognized that the "expertise, leadership, and resources of the NIH enable much rigorous science, interagency and intersector collaboration, and the public availability of biomarker data that would otherwise not occur." Furthermore, the report noted that the NIH "may also help play a role in prioritizing the development of biomarkers in underdeveloped areas, such as food and nutrition."

The workshop included two presentations from representatives of the NIH: Michael Lauer, director of the National Heart, Lung, and Blood Institute (NHLBI) Division of Cardiovascular Sciences, and Paul Coates, director of the NIH Office of Dietary Supplements. Dr. Lauer's talk offered a different perspective—one cautious about the utility of biomarkers—from that expressed in prior talks by representatives of the Food and Drug Administration's Center for Food Safety and Applied Nutrition (FDA/CFSAN) (see Chapter 3). John A. Wagner said that the committee did discuss perspectives on the general utility of biomarkers during their initial meetings. In the end, he said, "the committee did come to the conclusion that biomarkers are really an important part of medical research and medical practice, and we need to grapple with … the complex issues that

surround biomarkers, and the interfaces they have to other measurements in medicine."

The committee posed the following questions to the two speakers from the NIH, whose presentations are summarized below:

- How are biomarkers currently evaluated at the NIH?
- How does the NIH interact with the FDA regarding biomarker evaluation?
- Are there priority areas for research with respect to particular biomarkers?
- How does the recommended biomarker evaluation process differ from current processes at the NIH?

PRESENTATION BY MICHAEL LAUER, NHLBI

Dr. Lauer began by drawing an analogy between the use of biomarkers and an event in the history of baseball described in the book, *Moneyball: The Art of Winning an Unfair Game* (Lewis, 2003). To assemble a successful baseball team, Billy Beane, the general manager of the Oakland A's, was the first to apply rigorous analytic methods to identify which kinds of data best predict how much an individual player could contribute to a team's overall success, rather than rely on conventional metrics. Dr. Lauer noted that this strategy resembles the notion "that by using the right statistics—the right biomarkers, so to speak—like on-base percentage and pitch count, one can come up with winning baseball teams without having to spend that much money."

Turning to the report, Dr. Lauer noted that it identifies five types of biomarkers: physiological measurements, blood tests and other chemical analyses, genetic data, metabolic data, and measurements obtained from images. He provided illustrative examples—in the form of cautionary tales—representing each of these biomarker classes.

As an example of a physiological biomarker, Dr. Lauer described the exploration of ventricular premature beats, a form of cardiac arrhythmia. "In the 1980s, the presumed belief was that people who have lots of premature beats were at increased risk for [cardiac sudden] death, and if you got rid of the premature beats, you would save lives," he said. This hypothesis was tested in the Cardiac Arrhythmia Suppression Trial (CAST) (1989). So strong was the belief that suppressing premature beats was life-saving that some considered this trial to be unethical, he said. However, the results clearly associated treatment to suppress ventricular premature beats with increased death rates.

Decades later, the NHLBI funded another trial examining the potential of an implantable defibrillator to reduce sudden cardiac death rates,

said Dr. Lauer. This three-way trial randomized patients with heart failure to receive amiodarone, which at that time was believed to be the safest antiarrhythmia drug available, or a defibrillator (Bardy et al., 2005; Torpy et al., 2007). The result was that "amiodarone did nothing," he said. "It didn't make things worse, but it didn't make things better, either. But there was a reduction in death rate with the defibrillator." That finding has largely defined care for heart failure since then, he added.

Dr. Lauer's second case involved biomarkers measured by blood tests, an area that receives significant research funding from the NIH. He noted that in 2006, two NHLBI-funded studies were published that reached opposite conclusions. One, which originated from the Framingham Heart Study, examined a panel of biomarkers[1] in approximately 3,000 people and found that those who had more abnormal biomarker levels were found to have higher rates of major cardiovascular events (Wang et al., 2006). However, these findings were no more helpful than standard risk factors in identifying people likely to experience such events. The second study was a component of the Women's Health Study, and measured C-reactive protein (CRP) along with a variety of other biomarkers in 15,000 women (Cook et al., 2006). In this trial, CRP was found to improve risk discrimination as well as many other standard risk factors, said Dr. Lauer. However, the Women's Health Study data analysis differed from the Framingham study because it focused on a statistical approach known as reclassification, or the ability to reassign patients to different levels of risk. Cook and colleagues argued that this measure reflected the utility of CRP as a biomarker, according to Dr. Lauer.

For a genetic biomarker example, Dr. Lauer described a recent genomewide association study of Alzheimer's disease involving over 35,000 subjects, of whom 8,371 developed Alzheimer's disease (Seshadri et al., 2010). He noted that the researchers confirmed two loci associated with the disease, and that they had identified and replicated two more such loci, but that compared with three well-known risk factors—age, gender, and apolipoprotein E (APOE) levels—these additional genetic markers had absolutely no impact on disease risk. "The authors concluded that, while not clinically useful, the value of these genetic markers may lie in the insights they could provide for research into the pathophysiologic mechanisms of Alzheimer's disease," said Dr. Lauer. He further noted that Harold Varmus, who at the time was about to become the director of the National Cancer Institute, was quoted as saying "genomics and related disciplines are more closely aligned with modern science

[1] These included CRP, brain natriuretic peptide (BNP), aldosterone, fibrinogen, D-dimer, homocysteine, and the urinary albumin to creatinine ratio.

than with modern medicine. They produce knowledge that is broad in its scope, but only a few selected items of that new information are now widely used as guides to risk, diagnosis, or therapy" (Varmus, 2010).

As an example of metabolic data used as a biomarker, Dr. Lauer pointed out that it has been known for some time that elevated levels of hemoglobin A1c are associated with poor outcomes in patients with diabetes and more recently, in the general population. As a result, it was hypothesized that by aggressively reducing blood sugar, as reflected in reduced hemoglobin A1c levels, major cardiac events could be prevented. Dr. Lauer described the NHLBI-funded study, ACCORD (Action to Control Cardiovascular Risk in Diabetes), in which 10,251 patients were randomly assigned to intensive hemoglobin A1c reduction or to a more conservative hemoglobin A1c reduction strategy (Dluhy and McMahon, 2008). "The trialists were remarkably successful in getting the hemoglobin A1c down and keeping it down … but that did not translate into an improvement in outcome," he said. In fact, the trial had to be stopped prematurely because patients who were randomized to intensive therapy demonstrated an increased mortality rate. "These data have been analyzed over and over and over again, and to date it is not at all clear why this happened," he said.

The fifth and final biomarker Dr. Lauer described was an example of an imaging biomarker. In the 1990s, a series of observational studies found that patients who were clinically stable days after experiencing myocardial infarction (MI), and who went home with an open (known as "patent") artery resulting either from spontaneous fibrinolysis or from balloon angioplasty, had better outcomes than stable MI patients who went home with occluded arteries. These results became widely accepted as the "open artery hypothesis," said Dr. Lauer. There were some who questioned this theory, prompting the Occluded Artery Trial in which 2,166 patients were randomized to either receive a percutaneous coronary intervention (PCI)[2] or not (Hochman et al., 2006). "The results were a bit disappointing and surprising, in that the patients who were randomized to receive a PCI did not have an improved outcome, and if anything, it was a bit worse," he said. Within a year of this trial, the American College of Cardiology and American Heart Association changed their practice guidelines to indicate that PCI of a totally occluded artery after a completed infarct in an otherwise stable patient is not recommended (Antman et al., 2008).

Dr. Lauer then discussed NHLBI's approach to biomarkers, which he said includes biomarker discovery and the use of biomarkers to

[2] Also known as a stent.

- differentiate clinically relevant disease subtypes;
- identify new molecular targets for intervention;
- improve risk assessment, diagnosis, prognosis, and prediction of response to therapy; and
- develop personalized preventive and therapeutic regimens.

One example of NHLBI's current work in biomarkers is the Systems Approach to Biomarker Research in Cardiovascular Disease (SABRe in CVD) Initiative,[3] which is a plan to combine peripheral biomarkers, genetic markers, and markers of gene expression to create sophisticated models of clinical disease using the Framingham Heart Study and other patient cohorts. Dr. Lauer noted that the NHLBI is also involved in the Clarification of Optimal Coagulation through Genetics (COAG) trial,[4] which is a randomized clinical trial assessing the value of using genetic markers as a biomarker for patient response to anticoagulant therapy warfarin. Patients in the trial are being randomized to have their warfarin dosage determined by a standard clinical strategy versus a strategy informed by genomics; the primary outcome will be correct anticoagulation within a relatively short period of time, said Dr. Lauer. Another such study funded by the NHLBI, the Prospective Multicenter Imaging Study for Evaluation of Chest Pain (PROMISE) trial,[5] will randomize 10,000 patients with suspected coronary disease to an anatomic test, coronary computed tomography (CT) angiography, versus functional tests, including exercise electrocardiogram on a treadmill, stress echocardiography, and stress nuclear imaging. Dr. Lauer noted that in this study, the biomarkers themselves are being rigorously tested through a randomized clinical trial.

In concluding, Dr. Lauer noted that the report, which he called a "very thoughtful document," nonetheless contains some concerning statements. For example, the report states that biomarkers "can enable faster and more efficient clinical trials for lifesaving and health-promoting interventions." Dr. Lauer suggested perhaps the opposite of this statement is true, and reiterated hypothetical questions previously posed about hormone replacement therapy (HRT): what might have happened if, decades ago, a definitive trial was conducted to directly examine the effects of postmenopausal HRT on clinical outcomes instead of relying on a biomarker that was eventually disproven? How many women would not have been prescribed this therapy and how many lives could have been saved if a more expeditious course focusing on clinical outcomes had been taken (Califf, 2006)?

[3] See http://www.framinghamheartstudy.org/participants/news.html.
[4] See http://coagstudy.org/.
[5] See https://www.promisetrial.org/.

On the other hand, he applauded the report's statement that "we need to gather available evidence to predict effective interventions and clinical endpoints of interest," which, he said, "points out why we cannot avoid doing clinical endpoint randomized trials."

Dr. Lauer noted that years of research such as the studies he described, have led to a better understanding of the failings of biomarkers. However, this increase in knowledge means that it is now more difficult to make a general judgment about the value of all biomarkers.

PRESENTATION BY PAUL COATES, NIH OFFICE OF DIETARY SUPPLEMENTS

Created by an act of Congress in 1994, the NIH Office of Dietary Supplements (ODS) collaborates with other NIH institutes and centers on areas of common interest, including biomarkers, said Paul Coates. He used the example of vitamin D, a putative biomarker for various chronic conditions, to illustrate the challenges his office faces with respect to biomarker evaluation, and how they address these challenges through collaboration with a range of federal agencies.

Vitamin D status—measured by the level of 25-hydroxy vitamin D in blood—is most closely associated with rickets in children and osteoporosis in the elderly, said Dr. Coates. "Over the years, the very best data have been developed against those endpoints in those two rather specific populations," he observed, and the possibility that vitamin D status affects other chronic conditions is "part of a very active, ongoing collaborative set of studies that are funded by the NIH, with lots of input from others."

The ODS works with the Centers for Disease Control and Prevention (CDC) to enhance availability of data on vitamin D status in a nationally representative sample of the United States through the National Health and Nutrition Examination Survey (NHANES). Over the years that these measurements have been taken, several methodological challenges have emerged. For example, a proposed change in the method for measuring 25-hydroxy vitamin D revealed "major differences" in values measured by different methods, said Dr. Coates. To find a rational approach to deal with these variable measurements, the ODS, the National Center for Health Statistics (which runs the NHANES vitamin D study), and the National Center for Environmental Health (which measures 25-hydroxy-D levels in the NHANES samples), convened a roundtable discussion involving laboratory scientists, vitamin D researchers, statisticians, population scientists, and other experts, said Dr. Coates. The results of these discussions were published in the *Journal of Nutrition* (Yetley et al., 2010). Dr. Coates noted that this situation demonstrates the importance of the

analytic validation of biomarkers, and suggested that not enough attention is paid to this issue.

Dr. Coates then addressed the question of what vitamin D status actually indicates. For the purpose of argument, Dr. Coates suggested that the measured level of 25-hydroxy vitamin D in an average human derives from several sources, including sunlight, supplements, and foods. Whether vitamin D status indicates anything except level of risk for rickets in children and for falls and fractures in the elderly remains unknown. "Given that," he said, "there are important reasons for people to want to consider taking vitamin D: the notion that their risk for type 2 diabetes, type 1 diabetes, cardiovascular disease, cancer, and a great many other chronic disease endpoints may go down," so it is important that studies address the link between vitamin D and a variety of chronic diseases. He also noted that the Institute of Medicine is currently completing a review of dietary reference intakes for vitamin D and calcium. This report will be based on a systematic evaluation of the existing literature relating vitamin D status to chronic disease endpoints, other than falls and fractures and rickets. "In some respects this parallels the committee's analysis of biomarkers in chronic disease management" said Dr. Coates. "We'll see what that thoughtful approach to the available science means in terms of making recommendations for dietary intake."

Dr. Coates noted that the ODS convened a working group on vitamin D comprising representatives from the NIH institutes, the FDA, the United States Department of Agriculture (USDA), the CDC, and corresponding agencies in the Canadian federal government. Because the ODS is a small office with limited resources, its effectiveness results from collaborating closely with other agencies. He noted that this is an approach the ODS is also taking to study a range of nutrient biomarkers with potential relationships to chronic disease, including omega-3 fatty acids, folate, and vitamin B_{12}.

5

Industry Perspectives

Speakers representing companies and professional organizations with an interest in biomarkers gave 5-minute presentations and participated in a panel discussion. These individuals represented a range of industries, including food, grocery manufacturers, nutritional supplements, pharmaceuticals, medical diagnostics, and devices.

The committee asked these speakers to address the following questions:

- How do biomarkers impact patients, consumers, or clients in your industry stakeholder group?
- How does the recommended biomarker evaluation framework relate to biomarker evaluation processes currently in use with your industry or stakeholder?
- Will a prospective biomarker evaluation process add clarity to product development and consumer understanding?
- To the degree it is possible to respond, if at all, what are your key concerns for implementation of the recommendations?
- To the degree it is possible to respond, if at all, which recommendations of the report are the most useful or important?

Several themes emerged in the course of these presentations that were further explored during subsequent discussion. First, echoing its emphasis in the report, was the importance of context of biomarker use. As Allan Jaffe, professor of medicine at the Mayo Clinic, noted, these exchanges recapped discussions that occurred among committee members as they

determined whether to separate utilization from analytical validation and qualification in the recommended framework for biomarker evaluation.

On a related point, some discussants raised the concern that case studies in the report might be interpreted as questioning the value of useful biomarkers that are not appropriate surrogate endpoints (for example, C-reactive protein [CRP]). Guy Johnson remarked that the predictability and shortcomings of biomarkers such as prostate-specific antigen (PSA), CRP, and low-density lipoprotein cholesterol (LDL-C) are well known by medical professionals, and that this information should also be available to consumers. However, he said, "in a quick read of the report, I'm worried that its results could be misconstrued to undermine the value of some of these markers that are being used routinely."

Roberta Ness replied that the committee had shared this concern, and they included information and recommendations in the report emphasizing the importance of scientific literacy, numeracy, and communication with patients regarding biomarkers. "PSA is such a perfect example, because there is so much controversy about its use," she said, noting that physicians do not consistently interpret the results of this assay. Maria Lopes-Virella, professor of bioengineering at the Medical University of South Carolina, added that while providing information about biomarkers is important, the public needs to be made aware that the same biomarker can be valid for use in one context and not in another.

Dr. Jaffe noted that health care providers also need guidance on these issues. "Lots of physicians say, 'That's been in trial, [and] *X* showed better than *Y*; therefore, everybody gets *X*,'" he said. "That sort of thinking is a negative for the use of biomarkers, because it means when you use a biomarker, you [reduce it] to its lowest common denominator and overutilize it." Thus, it's important to encourage more probabilistic thinking on the part of physicians as well as the public to combat the misguided notion that a biomarker has to be good for every use or no good at all, said Dr. Jaffe.

Committee member Michelle A. Albert, assistant professor of medicine, Harvard Medical School, further noted that the phenomenon of "psychological toxicity" figured into the committee's deliberations. In the case of very ill patients, she said, it is important to consider what effect it may have to "use the shrinkage of a tumor … to tell someone that they are improving when in fact they are not, and … what that does to them psychologically as an important endpoint." That is an additional reason why the committee emphasized and separated context of use from biomarker validation and qualification, she stated.

Another concern raised by several discussants was the possibility that the standards established by the recommended biomarker evaluation framework will have a chilling effect on research in this area.

"I think it doesn't have a chilling effect at all," Jennifer Van Eyk responded. "I think it's very realistic, and I think people who are doing biomarker development—I can only speak from an academic [standpoint]—this is our reality. This is hard stuff to do. These are complicated diseases, and even when they are straightforward, the variation between patients, between individuals, between … population[s], is massive." However, researchers believe that it will eventually be possible to identify biomarkers as surrogate endpoints for chronic diseases. "We all realize it's a multistep pathway … [that] is well laid out [in the report]."

"I think we simply formalized what had been in the minds of many and maybe even in the operation of many," said Dr. Ness. The committee's recommended framework for all biomarkers reflects their concern regarding biomarker usage that puts large numbers of people at risk, especially when the risk of harm outweighs the potential benefit, she added.

Responding to this comment, Stephen Williams noted that biomarker qualification has typically been characterized by two competing philosophies. "One is to take the harm minimization approach, and to carefully characterize all the harm and to try and avoid it going forward," he said. The other philosophy is "the tolerability of risk approach, which is to carefully characterize the harm and to carefully characterize the benefit of going forward with a surrogate versus going forward with the best available alternative and to choose the better option." Both positions are expressed in different parts of the report, he observed, so he asked the committee, "is there an explicit embrace of tolerability of risk, or is there an explicit embrace of harm minimization? Or are people still ambiguous?"

John R. Ball answered that the committee tried to "point out the real benefits of biomarkers in a lot of cases" but also to express caution learned from experience. This was also their intention in recommending that the FDA assemble expert panels to evaluate biomarkers. Such panels would include people with explicit conflicts of interest because their expertise is needed. "That would be the sort of failsafe mechanism that would take into account the balancing of the benefits and the potential harms," he said.

David DeMets noted the contention—that few biomarkers could serve as surrogate endpoints—was stated more than a decade earlier (Fleming and DeMets, 1996). Not surprisingly, he observed, "most biomarkers put forth as surrogates have failed. By reflecting that reality, the report shouldn't chill biomarker research, he said, but focus efforts on more appropriate uses of biomarkers, such as the early screening of candidate drugs and the characterization of patient populations."

What really has a chilling effect, Dr. DeMets continued, is the inappropriate use of a biomarker. For example, in the early days of the HIV/AIDS epidemic, CD4 counts were used as a surrogate endpoint in drug

research. "Over time [the scientific community] realized they weren't such good surrogates," he said, but there were no alternative biomarkers for this purpose. Members of the HIV/AIDS patient advocacy community also questioned the use of CD4 counts as surrogate endpoints. "They demanded that we shape up and demand the best data, the best evidence," he said. "This is a population that had the most to lose, and yet they were the ones who understood what the risks were, because the [experimental AIDS] drugs were expensive, they were toxic, et cetera."

The distinction between foods and drugs, and in the ways each industry uses biomarkers, generated considerable discussion. Responding to Douglas Balentine, who remarked in his presentation (see below) that drugs cure and mitigate disease, while healthy foods reduce the risk of developing disease, Thomas Fleming noted that some drugs also serve the latter purpose; for example, interventions to prevent HIV transmission, type 2 diabetes, and Alzheimer's disease. Such drugs are rigorously scrutinized to ensure their efficacy, he continued, and if such claims are based on surrogate endpoints, those biomarkers must be validated. Since this is the case, he asked, "if foods want to have a health claim of reducing the incidence of cardiovascular disease or cancer, why shouldn't they undergo the same scrutiny as drugs would for making such claims?"

Dr. Balentine answered that conducting such prospective cohort studies for foods would be difficult because people who develop health problems are treated with drugs not food. While he agreed that rigorous science is necessary to support health claims for foods, he contended that a different approach is required as compared with drugs.

Dr. Krul noted that biomarkers currently used to make health claims for foods are relatively accurate and precise (for example, the 4 percent reduction of LDL-C claimed by Cheerios). The FDA wouldn't approve a new drug that lowered cholesterol by such a modest amount, she observed, "but is that good enough for a health claim?" And if such a claim is clinically relevant and statistically significant, she continued, does that make a food a drug?

Dr. Jaffe replied that the biological importance of a biomarker, such as LDL-C, should be considered during analytical validation and qualification. For example, he said, one could say that intravenous nitroglycerin following a mild acute infarction doubles the blood flow within the area of the infarction, but by increasing from 0.5 percent to 1.0 percent of normal flow, that gain is biologically insignificant. Proper analytical validation and qualification of biomarkers takes such information into account prior to considerations of context of use, he said.

Dr. Jaffe also noted that presenters from the food and nutritional supplement industries had expressed concerns as to how the recommended framework for biomarker evaluation would be implemented,

and in particular, how their industries could afford to conduct research on health claims. "I wish I knew all the answers about how to operationalize all of this in terms of … surveillance or appropriate studies," he said. However, he added, "one of the nice things about combining drugs, devices, nutraceuticals, and food under a similar umbrella is there is a huge amount of experience … at the industry level that can be tapped to help develop the appropriate paradigms." One such paradigm, Dr. Jaffe added, is likely to involve data sharing and collaboration on biomarker identification, as is already occurring among pharmaceutical and medical device companies.

PRESENTATION BY DOUGLAS BALENTINE, UNILEVER

Dr. Balentine said his remarks would try to illustrate the challenges and benefits of using biomarkers to support health claims in the food industry. "The industry needs clear guidance as our goal is to make truthful and not misleading claims and messages on products," he said. "As we develop products to help consumers choose healthier diets, clear science-based guidelines are essential." The food industry also needs to have a clear understanding of which biomarkers or surrogate endpoints the FDA will accept, based on rigorous scientific research, in support of health claims.

Foods and drugs are viewed differently, Dr. Balentine emphasized: foods can promote and maintain overall health and wellness while drugs can cure or mitigate disease. Both purposes demand scientific rigor, he said, but the type of studies required to demonstrate each type of claim would be different, as befits the context of use for a drug versus a food. However, studies of biomarkers that the National Institutes of Health (NIH) or pharmaceutical industry conduct can benefit the food industry, which can apply their results to help determine health claims for foods.

Drugs are often a specific compound that is used to produce specific effects, whereas foods consist of many compounds which may produce many different effects. The complexity of food makes the health effects associated with dietary interventions difficult to study, said Dr. Balentine. This problem is further complicated by the fact that people in control groups for food studies are often exposed to the components being tested (for example, vitamin C, carotenoids) through other aspects of their daily diet. Thus, he said, there cannot be a true placebo-controlled trial of most foods.

The food industry uses biomarkers to support health claims, which involve reduced risk for a chronic disease endpoint, and also to support structure–function claims, which involve maintainance of a healthy process. In either case, Dr. Balentine emphasized that "biomarkers need to be

strong and they need to be clear, and they need to be scientifically supported." The biomarker evaluation framework should help to establish that clarity, he added, provided foods, medical devices, and supplements are each evaluated in context and with appropriate scientific rigor.

PRESENTATION BY MELISSA MUSIKER, GROCERY MANUFACTURERS ASSOCIATION

The Grocery Manufacturers Association (GMA) represents the world's leading food, beverage, and consumer products companies, by promoting sound policy, championing initiatives that increase productivity and growth, and helping to ensure the safety and security of consumer packaged goods through scientific excellence, said Melissa Musiker, senior manager of science policy, nutrition, and health, at the Grocery Manufacturers Association.

She noted that the report's publication had sparked discussion in the popular press as to the validity and appropriateness of all types of voluntary nutrition claims used on food product labels. The GMA was concerned with this interpretation of the findings of the report because biomarkers are used in a different context when applied to food and voluntary nutrition label claims as compared with drugs. "Under the current regulatory framework, a variety of voluntary claims are permitted for use in food labeling, provided that they are truthful and not misleading," she explained.

"There are also three other types of voluntary claims related to nutrition and health that are permitted for use on food labels but without prior approval," Ms. Musiker continued. "The first is a dietary guidance statement, which typically refers to the benefits of the broad class of foods and reduced risk of disease or health conditions. The second is a structure–function claim … that encompasses both expressed and implied claims regarding the benefits of food or food components in promoting and maintaining normal structures and functions of the body. The third is a nutrient content claim, a descriptive statement that characterizes the level of nutrients in the food."

All three types of claims imply that consuming the food bearing them leads to the maintenance of health and promotion of normal physiology, Ms. Musiker said. "Any food or component of a food, characterized by a voluntary label claim, should be of importance in human nutrition by virtue of its presence or absence at the levels the claim describes." In fact, she added, "all information provided as a mandatory nutrition-labeling element within the nutrition facts panel is assumed to meet the same standard." Ultimately, she concluded, "all nutrients on the label should

be relevant to the promotion of good health and by extension have an impact on a biomarker."

"The GMA agrees with the committee that a consistent level of scientific rigor should be utilized when evaluating biomarkers and surrogate endpoints in their ability to serve as predictive models from morbidity and mortality outcomes," said Ms. Musiker. However, since health claims are preventive in nature, the type of study needed to assess such a claim should differ from that used to support a disease treatment, she argued.

"The scientific research needed to assess the validity of a biomarker for the purpose of voluntary nutrition claims presents unique challenges for foods," she said. "A large-scale, double-blind, placebo-controlled, crossover, randomized clinical trial would be very difficult to design and successfully implement when studying foods and diets," she noted. Echoing a point made by Dr. Balentine, she added, "when examining nutrients within a food as opposed to a nutrient consumed in isolation, it is very challenging to isolate the impact of just one nutrient or food, and then to generalize this finding across populations … because all people eat food as a part of a daily diet and a dietary pattern, which can be so highly variable."

All foods are composed of a number of components (for example, fats, carbohydrates, fiber, protein, water, micronutrients, and phytochemicals) known collectively as the food matrix, Ms. Musiker said. These complex and highly variable food matrices can modify nutrient bioavailability, such as when a nutrient is consumed in combination with other nutrients. Due to these modifying effects of food matrices and overall dietary patterns, the committee's recommendations, as she interpreted them, would require a unique study to determine the validity of a biomarker relative to a nutrition claim for any given food or combination of foods.

"It's really not practical for novel research to be required each time a company would like to put a nutrition claim on a food," she said. "There is a well-established body of science that provides the fundamental basis for the vast majority of voluntary nutrition claims, linking nutrients and foods in the diet to the maintenance of normal physiology and good health." She noted that most of this science is articulated in the *Dietary Guidelines for Americans* or the *Dietary Reference Intake* report (HHS, 2005; IOM, 2006).

"The GMA proposes that a system be developed to provide guidance to food companies when developing and applying voluntary nutrition claims," Ms. Musiker said. This system should be designed with the committee's recommendations in mind, but should also take into account the challenges presented by studying foods or nutrients consumed as components of a meal, within a diet, and as part of a person's lifestyle. "The GMA believes the types of scientific studies needed … for the evaluation

of the appropriateness of dietary guidance statements, structure–function claims, and even nutrient content claims, may be inherently different from those needed to develop health or drug claims," she added. Therefore, the GMA "encourages those evaluating nutrition label claims to remember that there is a distinction between health claims related to the prevention and mitigation of abnormal biomarker status and other types of voluntary claims related to the maintenance of normal status."

PRESENTATION BY ANDREW SHAO, COUNCIL FOR RESPONSIBLE NUTRITION

"The lack of validated biomarkers for exposure to nutrition interventions and surrogate endpoints for chronic disease limits the amount of research that can be conducted, especially for prospective randomized trials, due to cost and other logistical issues," Andrew Shao said. "This, in turn, limits the ability to derive answers to important questions relating to the ability of diet, food, and food components to promote health and reduce the risk of chronic disease." The Council for Responsible Nutrition (CRN) has commented on the need for additional biomarkers as surrogate endpoints to both the FDA and the NIH, he reported.

Having a framework for biomarker evaluation is a step in the right direction, Dr. Shao said. "We anticipate that a formal biomarker evaluation process will add clarity to product development, [and that] as companies that choose to invest in research [we] will have a better understanding *a priori* that the research will have broader acceptability and applicability to public health recommendations, such as health claims," he said. Regarding consumer understanding of biomarkers and their evaluation, he asserted, "in the end, consumers are not so much interested in the biomarker as they are in the clinical endpoint that it represents."

The CRN's primary concern regarding the committee's recommendations is that human or financial resources may prove inadequate to implement them, Dr. Shao said. He also expressed concern that some recommendations might be misinterpreted; for example, he interprets Recommendation 3 (see Box 2-1) to mean "when it comes to relying on a biomarker as a surrogate for a clinical endpoint, the product application, whether a food, drug, or device, is irrelevant, and that there should be a single standard," he said. He added that he agreed with this position, "whether we're talking about a statin or we're talking about dietary fiber."

However, he asserted, "text in the body of the report associated with the recommendation is not consistent with that interpretation." For example, he quoted from the report, "the FDA's regulation 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 nondrug products." That passage, he explained, caused him to ask whether the committee was suggesting that foods and supplements be regulated like drugs.

Calling Recommendations 1 and 2 (see Box 2-1) "the most useful and relevant recommendations," Dr. Shao said that they "represent what we believe the IOM committee was charged with accomplishing and set a solid foundation for a scientific framework that can be applied to the literature and future research efforts." On the other hand, he described Recommendations 3 through 5 as "both unnecessary and confusing."

"I certainly won't speak for the FDA, but it seems to me that recommendations dealing with food policy and regulation, including that of dietary supplements, seem to go beyond what the committee's mandate was," Dr. Shao said. "For example, Recommendation 4 really seems to relate to general issues regarding health claims … but it says nothing about biomarkers."

By contrast, "We feel Ancillary Recommendation 6a [which encourages the collection and sharing of biomarker data across the Department of Health and Human Services] is very important for the implementation of the first two recommendations." A number of different federal agencies conduct research on and evaluate biomarkers, Dr. Shao pointed out. "To fully leverage all of these resources, these efforts cannot be siloed, but instead should be consistently shared across the various agencies."

PRESENTATION BY STEPHEN WILLIAMS, SOMALOGIC, INC.

"Pharma is a very diverse range of interests," Dr. Williams pointed out, and while he did not seek to represent them all, he expected that his opinion, which reflects his many years of involvement in the industry, was "reasonably representative of mainstream opinion."

Dr. Williams said that in regard to biomarkers, the pharmaceutical industry wants the same thing as the FDA, as reflected in their charge to the committee: "to establish a framework for the level of evidence and the nature of evidence for biomarker utilization." However, such a framework should also enable drug companies and their regulators to reach a decision as to what level of evidence is needed for a particular biomarker application. He added that the pharmaceutical industry wants "consistency and transparency in biomarker evaluation, rather than something … unpredictable and opaque," as reflected in the report. By focusing on two questions, "how good is good enough?" and "what type of evidence?," he briefly reviewed how well he thought the report and its recommenda-

tions had answered these questions, and in what ways he felt they had fallen short.

The committee's recommendation that the FDA establish expert committees for biomarker evaluation across its regulatory areas "would move forward the ability to decide how good was good enough," Dr. Williams said. He similarly applauded the three-part framework, its consistent application across all biomarker uses, the continual reevaluation of biomarkers, and the comprehensive evaluation of risks associated with the use of biomarkers as surrogate endpoints. An especially important feature of the report is its advocacy for biomarker consortia, he added. Pharmaceutical companies had realized for some time that more evidence might be needed to evaluate a biomarker or a surrogate endpoint than any one firm could generate on its own, and that even if it were feasible, doing so would not often be cost-effective.

Turning to his concerns about the report and its recommendations, Dr. Williams reiterated a concern he had expressed earlier in the workshop, the notion that one must define a specific intended use for a biomarker in order to determine both the benefits of success and the consequences of error in its usage. "I don't think a committee's going to be sitting there wondering about what level of evidence there is for a biomarker when they don't have a purpose in mind [for it]," he said. Context "comes first because without that you cannot define an acceptable performance standard for analytic validation or qualification," he continued.

Dr. Williams also argued that the case studies in the report did not actually assess the recommended evaluation framework. For example, the report said that "tumor shrinkage is not acceptable as a surrogate endpoint … partly because the analytic validation is not very good." However, within individual clinical trials, Dr. Williams noted that tumor size can be gauged accurately, "so the analytic validation can be controlled." In addition, he suggested that tumor shrinkage could have been used as an example to demonstrate evidence thresholds. "I would accept completely that if I'm developing a drug … [that is] 14th in class, and the cancer [it treats] is already well served by lots of other drugs, then I'm not going to ask you to tolerate the risks of using tumor shrinkage," he said. "But if I'm developing a drug for a cancer where there is absolutely no therapy today, it is completely fatal, the drug is known to be safe, and if I delay that drug on the market by another year to … [use patient] survival [as the endpoint] … another 10,000 people will die … [then] the value of using tumor shrinkage in that case is 10,000 lives. The consequence of error, to say I'm wrong and the surrogate endpoint didn't work, is zero. It is the cost of the drug company expense on the drug, but actually, if those people were going to die anyway, letting the drug through until you find

out that it didn't translate into a benefit has no consequence at all," he asserted. "So in that case, I might choose to accept tumor shrinkage."

Regarding these points, Dr. Fleming later expressed concern that Dr. Williams was "advocating for a lower bar that could be used to justify [using tumor shrinkage] as a surrogate." Dr. Fleming noted the potential for false negative risks in such cases as renal cell cancer, which fit the scenario Dr. Williams described. There were no effective therapies for that cancer when sorafenib came along several years ago, said Dr. Fleming. If tumor shrinkage had been the measure by which that drug was judged, its benefit would have been missed because it doesn't shrink tumors; neither do some other drugs that have had major effects in colorectal and lung cancer, he added. He also noted the potential for false positive effects, which might arise when side effects significantly increase risk relative to benefit. Moreover, he added, the law states that approval of a drug requires evidence of efficacy, so standards for approval are not simply at the discretion of the FDA.

There are additional downsides to using an imperfect surrogate endpoint in such a situation, Dr. Fleming continued. A patient taking a therapy that doesn't work but was erroneously approved on the basis of a surrogate endpoint might otherwise have joined clinical trials for altruistic reasons, as well as for the potential benefit of the experimental therapy, he said. Finally, he argued, limited health care resources should be focused on therapies likely to provide more benefit than risk. "So," Dr. Fleming asked, "how can we defend a low bar for a surrogate that is very unlikely to reliably predict benefits simply because there are no other effective therapies?"

Dr. Williams agreed that there should be a "level playing field for benefit and risk." He said that he "wasn't advocating a blind acceptance of a surrogate just because there were no other therapies" but rather that its evaluation include an assessment of the consequences of all possible errors associated with its use, as compared to all possible associated benefits. Such an analysis might have excluded tumor shrinkage as a surrogate endpoint, he said, but the case study in the report should have reflected that process in its entirety. "What I was looking for was an example … that demonstrated to the audience how one might change the balance of the weight of evidence depending on the … harm of failure or value of success," he explained.

Dr. Fleming responded, "I believe the reason you didn't see that is when we have 35 years of experience using this measure and now have extensive data regarding the relationship between the effects on tumor shrinkage and the effects on clinical endpoints, we're finding that we're not seeing a reliable assessment of efficacy using tumor shrinkage."

But many readers of the report don't have the benefit of these years of experience, Dr. Williams pointed out, so they might have benefited from learning how the threshold for use was defined for tumor shrinkage as a surrogate endpoint, and how the biomarker doesn't meet this threshold. For example, the report might have stated how many instances when trials using tumor shrinkage as a surrogate endpoint had failed to translate into improvements in survival postmarket. Then, he added, the report could have explained why this statistic is not tolerable or under what circumstances it might be acceptable. "I was looking for some kind of connection between the failure rate of the surrogate and the value of using it versus the harm of not using it," Dr. Williams said.

Dr. Fleming maintained that it isn't appropriate to lower the bar for acceptability of surrogate endpoints when nothing else works, because "something unreliable could lead us to missing effective therapies or declaring that something should be used when it is actually unfavorable in benefit-to-risk."

In a similar vein, Dr. Williams said he felt the report underplayed the value of surrogate endpoints, particularly as compared with their risks. "The report really nicely goes through all the different kinds of risks that one might come across when using a surrogate endpoint," he said. However, if you don't recognize the outcome of a short trial in terms of lives saved, you wouldn't tolerate the risks inherent in using surrogate endpoints. Focusing on risk and downplaying benefit leads to "a kind of precautionary view of surrogate endpoints," much as Michael Lauer had expressed, Dr. Williams noted. "I'm surprised that someone from cardiovascular health can say that they are concerned about whether surrogates are useful, [given] the … millions of lives saved by using blood pressure or LDL [as biomarkers]," he said.

Another benefit of surrogate endpoints is the advantage they offer to researchers seeking new therapies, Dr. Williams said. He noted that a recent review found that diseases for which a surrogate endpoint or efficient clinical endpoint[1] is available have an average of 100 therapies on the market, as compared with an average of less than one therapy available for diseases that lack a surrogate endpoint or efficient clinical endpoint (Lathia et al., 2009). That's another reason to take a balanced view of tolerable risk in using surrogate endpoints, he suggested.

Dr. Williams was also concerned that the evaluation framework did not include an analysis of the cost-effectiveness of surrogate endpoints. He argued that cost-effectiveness is a standard of the British National

[1] An efficient clinical endpoint is defined in Lathia et al. (2009) as those that enable proof of concept studies in 6 weeks or less and pivotal trials in 6 months of dosing or less.

Health Service, the largest single-payer health care organization in the world, and is also a factor in decision making for many federal agencies, including the Environmental Protection Agency and the Federal Aviation Administration.

Finally, Dr. Williams noted that pharmaceutical companies have been involved in the testing and validation of biomarkers for at least a decade, and that some people within the industry have run dozens or even hundreds of biomarker validation trials. "If there is to be a committee in the future that makes recommendations on whether … [a biomarker or a surrogate endpoint] is acceptable or not, the conventional wisdom would say, don't include industry experts, because they are conflicted," Dr. Williams said. He would prefer to see people from industry contribute their considerable expertise to biomarker evaluation committees. Responding to this point in discussion, Dr. Ball noted that the committee recommended that such expert panels comprise people representing a range of perspectives, because those are the people with the greatest expertise on the topic; he also noted that the FDA typically includes experts with these types of conflicts of interest on such panels.

PRESENTATION BY JAMES MAYNE, PFIZER, INC.

At the outset of his remarks, James Mayne, senior director at Pfizer congratulated the committee on its work: "This is clearly a very timely, very powerful, and very scholarly work that sets a framework and a roadmap for the future of biomarker development and qualification." Speaking on behalf of Pfizer and himself, Dr. Mayne noted, "we are very, very appreciative that the committee didn't step back from the ambitious scope that it undertook at the outset of this work." He believed the committee has created a unified decision-making system that encompasses all types of biomarkers, products, and decisions regarding their usage.

Recalling a presentation he delivered to the committee during its deliberations, Dr. Mayne noted that he had advocated a two-step framework consisting of analytical validation for biomarkers, followed by "fit-for-purpose" qualification. When he first reviewed the report, he said, he questioned why the qualification and utilization were not integrated. However, "as I read through the report and saw the rationale, it does make a tremendous amount of sense; … [the framework] can be used across many different product types and many different market types."

The report clearly defines what the framework is not intended to shape, Dr. Mayne emphasized. "It is not intended for biomarkers used in the discovery space," he said. "It is not intended for the 'tools of the trade,' as we sometimes call them in the drug industry, by which we make decisions on which chemical moieties to advance, which diseases to pursue, and where

the best opportunities may lie. Very wisely … the committee made it clear that those areas are omitted so as to not constrain innovation."

At the conclusion of his remarks, Dr. Mayne offered what he called "a criticism and a challenge." He said that the report failed to specify the "actual elements of the decision framework," or the criteria by which regulatory decisions should be made. "That was not provided, at least not in the detail I was looking for," he said. "That's fine; you have to start somewhere."

PRESENTATION BY JACK ZAKOWSKI, BECKMAN COULTER, INC.

Although he noted that his comments did not necessarily represent the position of his company, nor of the diagnostic industry in general, Jack Zakowski, director of scientific affairs and professional relations at Beckman Coulter, Inc., began his remarks by voicing general agreement with the report's evaluation framework and in particular, the interdependence of its steps, as depicted in Figure 2-1. However, he thought the arrows in the figure should be double-headed, indicating bidirectionality, to better reflect that interdependency.

Dr. Zakowski then sought to answer the session questions, as stated above. He said *in vitro* devices (IVD) that assay biomarkers aid in the overall assessment of patient status by diagnosing and monitoring chronic disease, guiding therapy, and predicting outcomes. The committee's framework resembles current practices of biomarker evaluation within his industry, but he emphasized that IVD assays measure biomarker concentrations, and are not biomarkers in and of themselves. "I think too often we have not drawn that distinction," he said.

In considering whether a prospective biomarker evaluation process would add clarity to product development or consumer understanding with respect to IVD biomarker assays, Dr. Zakowski noted that such clarity is achieved through the following routes, most of which were also noted by Dr. Williams:

- Clear definitions of terminology,
- Common conceptual framework (as shown in Figure 2-1),
- Clear common goals,
- *A priori* assessment criteria and analysis tools, and
- Transparency and predictability of process.

Regarding terminology, he asked, "when we use words as supposedly simple as *sensitivity* and *specificity*, are we talking about clinical or diagnostic sensitivity and specificity, ROC [receiver operating characteristic] curves, or what? Or do we mean analytical sensitivity, which could be a

lower limit of detection? Or freedom from interferences?" Such clarity is needed in the report, he observed.

Dr. Zakowski listed three key concerns for implementing the committee's recommendations: the potential conflation of assays with biomarkers (as previously stated); the need to emphasize that assays are both independent of and interdependent on the biomarker's utility; and the tendency toward "on-off" clinical thinking that ignores the fact that biomarkers measure the likelihood of disease and not disease itself. He described several examples of tests that have been improved by the selection of a better biomarker, such as the transition from glycated hemoglobin to hemoglobin A1c as an indicator for diabetes risk. He also noted that some assays have evolved to better measure biomarkers, as occurred with the advent of ultrasensitive thyroid-stimulating hormone (TSH) tests.

Concerns that biomarker assays encourage "all or nothing" thinking are another way of saying "context matters," Dr. Zakowski observed. The upper limit of normal "cholesterol … is 200," he said. "If your cholesterol is 199, you are not going to get prescribed a statin. If your cholesterol is 201, you are going to get prescribed a statin," he said, because "medicine is practiced like that too often." Thus, as biomarker assays improve, care must be taken to assess their results in the context of everything known about each patient, he warned. "Don't allow a specific number to drive a patient diagnosis, or treatment, or therapy."

In concluding his talk, Dr. Zakowski observed the following aspects of the report as being the most useful and important:

- Provide clarity and common understanding of definitions, process, and criteria.
- Provide transparent scientific basis for evaluation.
- Recognize the interdependence of analytical validation, qualification, and utilization.

PRESENTATION BY RICHARD KUNTZ, MEDTRONIC

Richard Kuntz, senior vice president and chief scientific, clinical, and regulatory officer at Medtronic, Inc., noted that his remarks reflected a quick review of the report from his personal perspective and not as a representative of the medical device industry. He began by describing the approval process for medical devices, which commences with a premarket or preapproval phase characterized by "a lot of complicated bench measurement," he said. He noted efforts underway to reduce the amount of testing and timing that occur during this phase in order to expedite the device approval process through such means as computational bioengineering modeling.

Device failure—especially for implantable preventative devices such as cardioverter-defibrillators, deep brain stimulators, or insulin infusion pumps—does not manifest as a clinical endpoint, making surrogate endpoints for the failure of these devices desirable, Dr. Kuntz said. However, these are difficult to develop. He said that his own efforts to design a model to measure the narrowing of a coronary artery—which he characterized as a "very simple process … and mechanism"—was a difficult endeavor with important caveats and limitations (Mauri et al., 2005).

"We started out many, many years ago trying to describe the dynamics of opening and closing an artery by a stent or any other device by using metrics of diameter measurements, which we called 'acute gain' and 'late loss,'" Dr. Kuntz said. "Our interest was to measure this late loss, which we could do with some degree of accuracy through quantitative angiography as a surrogate … for the need for revascularization, which was the clinical interest," he continued. To do this, he and his colleagues examined the results of several studies, including both drug-eluting stents and bare metal stents for which the endpoint of interest was target lesion revascularization (TLR) rate or need for revascularization, he said.

The researchers found a loose correlation between late loss and TLR, after which they conducted a series of studies in an attempt to describe this relationship in a sufficiently reliable way so that TLR could serve as a surrogate endpoint for late loss, said Dr. Kuntz. They also found that perturbing this system with treatments resulted in correlated responses by TLR and late loss. After more than 15 publications on their methodology over the course of nearly a decade, he and his colleagues felt they had a established a simple relationship between late loss and TLR and a relatively simple mechanical model explaining the relationship, which would be useful in studying late loss (Mauri et al., 2005).

However, Dr. Kuntz noted, as a surrogate for coronary stents, their model was ultimately limited. "I think we demonstrated through a lot of work that it was a good surrogate for binary angiographic restenosis and probably TLR," he said. However, in addition to revascularization—which is the main interest of reducing restenosis with such devices as drug-eluting stents—it is also important to measure stent thrombosis of other revascularizations outside the artery, he said. Their model did not account for such effects, "so therefore, it really did not turn out to be a fantastic way" to test new devices.

"We really do have few device surrogates at this point, and they certainly are not as simple as the one we studied," Dr. Kuntz summarized. "I think the biggest interest we have is trying to allow the timing of postapproval to occur during the product life cycle in time that can keep pace with technology, and focus more on the postmarket analysis," he said. Dr. Kuntz expressed his agreement with Recommendations 5 and 6. "There has to be

more rigor in the postmarket, especially for devices," he stressed. "We have to be able to look at product release, from efficacy studies to the real world, in a variety of dimensions … [involving] new operators and new patients, with more efficient systems in the postmarket." He added that the development of patient registries, along with advances in computational models and observational statistics, should encourage such efforts.

The concept of surrogate endpoints could be extended to several device design elements such as computational bioengineering modeling, Dr. Kuntz suggested. "Product performance is another interesting endpoint in and of itself," he added, which could be measured when a product that has no surrogate fails. Such an event could also be recorded in a patient registry, he said.

In general, Dr. Kuntz reflected that more rigorous postapproval studies need to be conducted, using better observational statistical methods, in order to balance pre- and postapproval data collection and to keep pace with rapid technological development.

6

Public Health, Consumer, and Consulting Organization Perspectives

Speakers from the American Dietetic Association (ADA), the Center for Science in the Public Interest (CSPI), and Johnson Nutrition Solutions, LLC, discussed their views on the committee's recommendations and report. These speakers were also asked to answer the same series of questions posed to representatives from stakeholder industries:

- How do biomarkers affect patients, consumers, or clients in your industry stakeholder group?
- How does the recommended biomarker evaluation framework relate to biomarker evaluation processes currently in use with your industry or stakeholder?
- Will a prospective biomarker evaluation process add clarity to product development and consumer understanding?
- To the degree it is possible to respond, if at all, what are your key concerns for implementation of the recommendations?
- To the degree it is possible to respond, if at all, which recommendations of the report are the most useful or important?

PRESENTATION BY GUY JOHNSON, JOHNSON NUTRITION SOLUTIONS, LLC

Guy Johnson said that his remarks did not represent any specific group but derived from his extensive experience with biomarkers as applied in such contexts as health claim substantiation.

He began by stating that the report outlines clearly "all of the factors and potential pitfalls" involved in determining what he called a "rock star biomarker." Calling the three-step framework comprehensive, Dr. Johnson agreed that biomarker use demands analytical rigor, and he assessed the qualification phase as feasible based on prospective epidemiological studies. However, he added, "the rub comes in determining the clinical outcome of an intervention where you change just the biomarker and nothing else, and then wait to see what happens in a group of healthy people." The report conveys the message that there are no shortcuts to a surrogate endpoint, he said.

Nevertheless, Dr. Johnson emphasized that biomarkers offer an important means to communicate useful information to companies and other organizations and to inform research. He expressed hope that the report can be used to "provide some structure around how that information can be communicated." Such communication could happen during the utilization step, by distinguishing between predictive biomarkers, for which there is evidence of an association with a clinical endpoint (for example, C-reactive protein [CRP]) that could be used to inform less rigorous claims, from probable biomarkers for which validity is clearly established. "It is possible that there could be some kind of regulatory language that would allow information on those less-than-rock-star biomarkers to be communicated," he suggested. "I don't think the committee's intent was to tell clinicians they can't measure CRP because it hasn't been fully established as a biomarker," he said; rather, there need to be guidelines for how to make use of information from biomarkers that are not surrogate endpoints. "That's where I think the real opportunity of this report is," he said.

Having the same standards of scientific rigor for biomarkers used in foods and drugs makes perfect sense, Dr. Johnson said. However, there are inherent differences between the process one might use to determine the effect of a food on a biomarker and the effect of a drug on the same biomarker. For example, he continued, "it's tough to do a randomized double-blind placebo-controlled trial with pomegranate juice [so that] … people don't know that they are eating it."

The report presents an opportunity to communicate information to consumers, as do biomarkers themselves, Dr. Johnson said. "Frankly, there's a danger in not communicating information to consumers," he said. On the other hand, he hoped that the report was not viewed as the "kiss of death" for a variety of biomarkers, such as low-density lipoprotein cholesterol (LDL-C).

Dr. Johnson also asked, in light of the report's recommendations, whether the Food and Drug Administration (FDA) would need to convene a panel to examine the blood pressure/sodium biomarker. "There are no clinical trials that show that reducing sodium, and therefore blood

pressure, in the population results in a clinical benefit," he said, "and there are suggestions that there could be unanticipated consequences [of sodium level reduction]. So if you apply the standards in this report, somebody has to take a look at that in a very rigorous way."

PRESENTATION BY MARY HAGER, AMERICAN DIETETIC ASSOCIATION

Mary Hager, director of regulatory affairs at the ADA, presented the association's viewpoints on the report. She noted that the ADA has more than 70,000 members, the majority of whom are registered dieticians who work in facilities and outpatient care. Biomarkers are part of the process by which these professionals make recommendations for changes in dietary regimens. In addition, registered dieticians also conduct research on nutrition, provide consumer education, and monitor the nation's health through such programs as the NHANES. The ADA is known for its innovative evidence analysis library and for its systematic review of the literature, the results of which are archived transparently, she said.

Nutritional intake and health status relate to chronic disease in terms of risk and disease progression, Dr. Hager stated. The ADA "recognizes that validated biomarkers are important to identifying estimates of optimal dietary nutrient needs and are important in determining dietary reference intakes," she said. In the past, when food was in shorter supply, research was conducted on the effects of an exposure to a nutrient in hopes of eliminating deficiency diseases, she explained. The abundance of calories in the current American diet does not guarantee nutrient sufficiency, let alone the optimal or ideal intake of particular nutrients, she said.

"Biomarkers of nutritional status as they relate to health are very important, and understanding the interplay between the nutrient intake and the development and progression of these chronic diseases is very important for individuals and consumers at large," Dr. Hager emphasized. For these reasons, the ADA supports the committee's recommendation that the FDA employ the same degree of scientific rigor in evaluating biomarkers across all of its regulatory areas. "Knowing with confidence the effect of a nutrient on an individual's or a population's risk for a disease or nutrition disorder would allow for more effective interventions," she said.

Dr. Hager also raised a topic of concern for the ADA regarding the use of genetic biomarkers in tests that she said she believed "are probably not approved or reviewed, and are … being used by individuals and by consumers … to diagnose allergies or disease." This phenomenon concerns many health care practitioners who are ADA members, she said, and expressed hope that the FDA "can have the authority to vali-

date and enforce surveillance of these … so-called diagnostic kits in the marketplace."

"I believe, and so do most of our members, that consumers will greatly benefit from the FDA having improved authority to test and to require valid data to support health claims, resulting in enhanced consumer confidence in selecting products that are best suited for their individual needs," Dr. Hager said. Regarding the committee's recommendations to this effect, she (like several other speakers) expressed concern that the necessary research would be difficult to fund, given thin profit margins associated with nutrients and foods as compared with pharmaceuticals. She also noted that the FDA will have to obtain "additional legal authority to request studies and sufficient authority to act on the results of the stud[ies] related to consumer understanding of claims on foods and dietary supplements."

PRESENTATION BY ILENE HELLER, CENTER FOR SCIENCE IN THE PUBLIC INTEREST

Ilene Heller, senior staff attorney at the Center for Science in the Public Interest, emphasized the importance of scientifically valid information on food labels that allow consumers to choose healthy foods and reduce their risk of diet-related diseases. According to Ms. Heller, the lack of consistent, reliable information on food labels necessitates more stringent standards for all health-related claims on foods. "For consumers, the most important recommendations to come out of the … report are the need for all health-related claims to be based on stringent standards and allowing claims to appear only on products that are part of a healthy diet," she observed.

Ms. Heller described three categories of health-related claims: authorized health claims, which require significant scientific agreement; qualified health claims, which are based on credible evidence; and structure–function claims, which merely associate a nutrient contained in the food with a health benefit.

Health claims are the gold standard, Ms. Heller said. "They require notice and comment rulemaking," she noted, and the foods to which they apply must contain a minimum level of nutrients prior to fortification (which is known as the "Jelly Bean Rule," because you can't fortify a jelly bean and call it healthy). Foods with health claims cannot exceed disqualifying levels for nutrients that need to be limited, such as sodium, cholesterol, or fat, and the health claims "must state that a disease depends on many factors, and that the nutrient must be eaten as part of a healthy diet." Some health claims state how much nutrient is needed to achieve the purported benefit, as well as how much of that nutrient is present in the food.

According to Ms. Heller, a problem that is occurring with increasing frequency—as typified by the Cheerios box that figured prominently in workshop discussion—is that health claims are being overshadowed by lesser claims (for example structure–function and qualified health claims) displayed in large print on the front of packages.

Since 2003, qualified health claims—a category originally applied only to dietary supplements—has been used for foods that attain the standard of credible evidence, said Ms. Heller. "Qualified health claims can even be made when the FDA thinks such claims are unlikely to be valid," she noted. For example, she quoted the language of the qualified health claim for green tea and cancer that was accepted by the Center for Food Safety and Applied Nutrition (CFSAN): "Two studies do not show that drinking green tea reduces the risk of breast cancer in women, but one weaker, more limited study suggests that drinking green tea may reduce this risk. Based on these studies, FDA concludes that it is highly unlikely that green tea reduces the risk of breast cancer" (FDA, 2009).

Qualified health claims are "pretty clumsy" and therefore not widely used on food labels—except for nearly half of all nuts—but they do "stimulate misleading advertising and public relations campaigns," said Ms. Heller. She also noted the results of two FDA studies conducted in 2005 and 2007 (which CSPI obtained by filing a Freedom of Information Act petition) suggested that consumers cannot distinguish between qualified health claims and "gold standard" health claims.

The most prolific type of health-related claim is the structure–function claim, Ms. Heller said. "That is not surprising … it is short and sweet, and it doesn't mention a disease. It is a lot more interesting for consumers and a lot more pleasant to see it on a cereal box than talk of cancer and heart disease." Moreover, no prior approval is needed to make such claims, nor do they require minimum or maximum nutrient levels. Often, if a company cannot make a qualified health claim for a product because it has too much fat, they make a structure–function claim instead," she observed.

While structure–function claims for foods cannot be false or misleading, the FDA has admitted to the Government Accountability Office (GAO) that the agency can't prosecute breaches of structure–function claims because it lacks resources for conducting research necessary to mount such a legal challenge (GAO, 2000), Ms. Heller said. However, the City of San Francisco successfully challenged a false immunity claim displayed on Cocoa Krispies boxes, as did the Federal Trade Commission, while the FDA did not. Recently, the FDA sent a warning letter to the makers of Juicy Juice regarding its structure–function claim linking docosahexaenoic acid (DHA, an omega-3 fatty acid) with brain health, she said, but only on the grounds that such claims cannot be made on products marketed to children under 2 years of age.

CSPI's conclusions regarding these matters are simple, Ms. Heller said: "Consumers can't tell one type of claim from another. Therefore, the same evidentiary and eligibility standards should apply to all categories of claims." To those who would argue that it would be too difficult to change the standards for all claims, she insisted this was not the case. "It is not going to be burdensome at all," she said. "There are well-established claims, like 'calcium builds strong bones,' that could form a category of claims to be included in a safe harbor or a positive list, as it is known in the European Union." In addition, health claims could be rephrased as structure–function claims, provided structure–function claims meet the minimum and maximum nutrient levels currently demanded by health claims. The FDA could also consider the structure–function claims approved by the European Food Safety Authority, she added. While CSPI delivered these recommendations to the FDA in 2008 and 2009, she said, the agency has yet to take action on them. CSPI's recommendations are consistent with the committee's Recommendations 3 and 4, Ms. Heller noted, "because they are based on the premise that all claims should be based on stringent scientific standards and should only appear on products that are part of a healthful diet."

7

Presentation by Thomas Fleming: Biomarkers and Surrogate Endpoints in Chronic Disease

Thomas Fleming provided a keynote presentation on the critical issues involved in the validation of surrogate endpoints. In his introduction of the speaker, John R. Ball noted that Dr. Fleming's work, and in particular, his publication with David DeMets (Fleming and DeMets, 1996), was influential to the committee and its recommendations.

Dr. Fleming began his talk by returning to a topic raised in discussion following presentations by stakeholders from industry: Is the report going to have a chilling effect on biomarker research and application? In his experience, such problems have resulted from "not having some consensus, both from a regulatory and scientific perspective, as to what it is that we need to show." Therefore, he believes the report should actually counter chilling effects. More importantly, interest in the report should be focused on asking whether it offers "enlightenment about how to enhance providing the public an informed choice, as well as more enhanced, reliable evidence about benefit to risk." In that regard, he said, he was very impressed with what the report had achieved "in taking on this complicated set of issues."

A CORRELATE DOES NOT A SURROGATE MAKE

Dr. Fleming focused on two main issues, the first of which he described as "digging deeper" into the reasoning behind his statement that "a correlate does not a surrogate make." He identified three main criteria for choosing endpoints in clinical trials in order to best determine

an intervention's benefit relative to its risk: sensitivity, measurability/interpretability, and clinical relevance. He illustrated the first criterion, sensitivity, with the choice of an endpoint for the trial of an analgesic for pain in preterminal cancer patients. Survival is critically important to such patients, he noted, but pain relief is the most sensitive measure of efficacy for this intervention.

The second criterion involves both measurability and interpretability. Dr. Fleming illustrated poor measurability with a hypothetical study requiring monthly liver biopsies. If such a study were conducted, it would not be acceptable to many patients and clinicians, and "you're not going to retain patients very long." He noted that interpretability is also important in selecting endpoints in clinical trials, and he provided some thoughts on composites of disease. One composite of disease, the combination of cardiovascular death, stroke, and myocardial infarction (MI), is interpretable because each of these conditions results in irreversible morbidity and mortality. However, interpretation becomes much more complicated if putative surrogate elements, such as "asymptomatic distal deep vein thrombosis," are added to the composite (as they often are in studies of knee or hip replacement, he noted).

Clinical relevance of the endpoint is the ultimate criterion for its acceptance, according to Dr. Fleming. He referenced Robert Temple's definition of a clinical endpoint: "a direct measure of how a patient functions, feels, and survives," which is also reflected in the Biomarkers Definitions Working Group and the Institute of Medicine (IOM) committee's reports (Biomarkers Definitions Working Group, 2001; IOM, 2010). Each of these attributes are difficult to measure, he acknowledged. Survival takes a long time to assess in many settings, and patient feelings and function are often based on patient-reported outcomes (PROs), which "can be very difficult to validate, often can have missing data, require blinding, and have a multiplicity associated with them," he said. "It's very tempting to look at objective alternative biomarkers."

A common approach to finding such biomarkers is to identify one that is correlated with the desired clinical endpoint, show an effect in the biomarker, and make the "leap of faith" that this biomarker does, in fact, translate to clinical benefit, Dr. Fleming said. "Unfortunately," he added, "that's often not the case." He proceeded to discuss the various reasons why a biomarker might fail as a surrogate endpoint, each of which are illustrated diagrammatically in Figure 7-1.

The first reason for a biomarker to fail as a surrogate endpoint is that the biomarker does not lie in the causal pathway by which the disease influences the clinical endpoint, so an intervention's effect on the biomarker will not provide a reliable estimate of the intervention's clinical efficacy (as shown in Figure 7-1A), Dr. Fleming said. An example of this

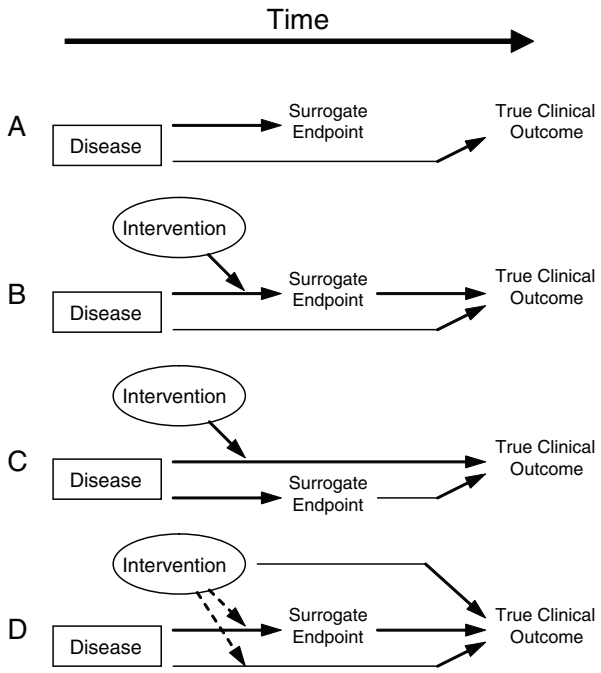

FIGURE 7-1 Reasons for failure of biomarker (surrogate) endpoints. (A) The surrogate is not in the causal pathway of the disease process. (B) Of several causal pathways of disease, the intervention affects only the pathway mediated through the surrogate. (C) The surrogate is not in the pathway of the intervention's effect or is insensitive to its effect. (D) The intervention has mechanisms of action independent of the disease process.
NOTE: Dotted lines = mechanisms of action that might exist.
SOURCE: Fleming and DeMets, 1996. Adapted, with permission, from the *Annals of Internal Medicine*. Copyright 1996 by the American College of Physicians.

case occurs in mother-to-child transmission of HIV. In pregnant women with HIV, there is a very strong negative correlation between maternal CD4 (helper T-cell) count and likelihood of HIV transmission to the infant, he said. Thus, one might suppose it would be useful to give the mother interleukin-2 (IL2) late in pregnancy in order to raise her CD4 count closer to normal levels. However, doing so has no effect on transmission because CD4 count is not part of the causal mechanism for HIV transmission.

Many diseases have biomarkers that are not in their pathophysiologic pathways, Dr. Fleming noted. Examples include carcinoembryonic antigen, a biomarker for ovarian cancer, and prostate-specific antigen (PSA); such biomarkers are useful for disease diagnosis and assessing prognosis,

he said. However, in the case of prostate cancer, he noted that controversy has arisen as to whether PSA levels should dictate the type of intervention undertaken, and also as to the value of ongoing assessment of prognosis for this typically latent disease. Correlation with a clinical endpoint is all that is needed for a biomarker to be useful in detecting disease and assessing prognosis, Dr. Fleming concluded. However, a biomarker used as a surrogate endpoint must lie in the pathophysiologic causal pathway of the disease.

A second scenario for biomarker failure as a surrogate endpoint occurs when multiple pathways influence outcome. If the intervention only affects the disease pathway through the biomarker (Figure 7-1B), a false positive can result, Dr. Fleming said. Conversely, if the intervention only affects pathways other than the one including the biomarker, the result would be a false negative (Figure 7-1C).

An example of the latter case involved chronic granulomatous disease—a condition that occurs in children who have a compromised immune system, resulting in serious infections, according to Dr. Fleming. Researchers considered treating this disease with gamma-interferon to increase bacterial killing. In fact, he said, bacterial killing was going to be used as the endpoint, "because we did not want to randomize and expose half of these children to three injections of placebo per week to gamma-interferon." However, due to concerns this surrogate endpoint might be misleading, a 12-month trial with interim analyses was conducted (Gallin et al., 1991). Early results of this trial were persuasive, yielding a 70 percent reduction in the clinical endpoint, but no effect at all on the biomarker, Dr. Fleming said. Gamma-interferon did indeed provide clinical benefit to affected children, perhaps by killing bacteria at an undetectable level, or through some other means, such as by increasing antibiotic uptake, he speculated.

Another example involves vancomycin-resistant enterococci (VRE) infections in the gastrointestinal (GI) tract. Patients with these infections are at considerable risk for bloodstream infections due to VRE, said Dr. Fleming. However, because more than 1,000 patients would be required to test the effect of an antimicrobial on that endpoint, researchers were interested in using decolonization of the GI tract by VRE as a biomarker for infection clearance. He noted that this biomarker has several flaws: it fails to take into account VRE colonization outside the GI tract, such as in or on the skin, and also for the magnitude and duration of antibiotic effect needed for protection. "If VRE GI levels are reduced to lower than detectable levels, it doesn't mean eradication," he said. "It's entirely possible that … you would still have risk of bloodstream infections." He also said that the method of quantification of VRE from fecal samples may not have adequately captured the extent of colonization. In addition to uncertain-

ties regarding the magnitude of effect on this biomarker that is needed for protection, there also is lack of clarity regarding how long VRE has to be cleared from the GI tract in order to reduce risk. The VRE decolonization biomarker also is unable to capture unintended effects of antimicrobials, such as suppressing the immune system or altering the composition of the GI flora, both of which could lead to opportunistic infections by other microbes. Indeed, he said, "there are many aspects of the ultimate effect of antimicrobials on the clinical endpoint that may not be captured by the biomarker [VRE decolonization]."

Even for interventions that affect all causal pathways leading to a clinical endpoint, there are potential off-target effects through which the intervention can directly influence the true clinical endpoint, and which the biomarker does not capture (Figure 7-1D), according to Dr. Fleming. One such scenario is the suppression of cardiac arrhythmia post-MI with ecainide or flecainide, which in a placebo-controlled trial were shown to triple the death rate among patients (Bigger, 1986; Cardiac Arrhythmia Suppression Trial [CAST] Investigators, 1989; Echt et al., 1991; Mukharji et al., 1984; Ruberman et al., 1977). There are many more such examples among treatments for cardiac arrhythmia (for example, quinidine, lidocaine) and among agents that improve cardiac output and ejection fraction (for example, milrinone, flosequinan), all of which were found to increase patient death rates.

Regarding cholesterol, Dr. Fleming noted that a meta-analysis of 50 trials of early generation agents for lipid-lowering—diet, lovastatin, fibrates, resins, and hormones—found that these agents produced a 10 percent reduction in low-density lipoprotein cholesterol (LDL-C), but no overall impact on overall survival (coronary heart disease [CHD]-related death was reduced, but non-CHD related deaths were increased) (Gordon, 1994). Later, the more potent statins produced a 30 percent reduction in LDL-C, and researchers began to see a relationship where such reduction in LDL-C predicted an effect on overall survival, he said. However, "as we've heard, when we then introduced torcetrapib with atorvastatin to achieve increases in HDL-C [high-density lipoprotein cholesterol] as well as reductions in LDL-C, this study was terminated early, surprisingly, with increased death and increased rates of CHD death, MI, and stroke" (Barter et al., 2007).

These results raise the important issue of bridging, Dr. Fleming said. This occurs when a biomarker that is a valid surrogate for a clinical endpoint—as effects on lipids were for the effects of atorvastatin on death, stroke, and MI—is proposed as a surrogate endpoint for a new intervention with a potentially different mode of action (for example, torcetrapib). In that case, the effect of that new intervention on the surrogate may not reliably predict the clinical endpoint, he emphasized.

Paula Trumbo later asked whether this concern applied equally to

surrogate endpoints validated for drugs that might be used to support a health claim for a food. Dr. Fleming constructed such a scenario—involving blood pressure effects that had been validated in a hypertensive setting for several classes of drugs—proposed as a biomarker for foods. "If I had my ideal, I'd like to validate each biomarker for classes of agents," he said, adding that he might be more willing to accept such an extrapolation if the same putative mechanisms were involved in both food and drug effects. One could say that foods are less likely to produce adverse (off-target) effects than drugs, he continued, but one could also argue that drugs might be more likely than foods to have broad positive (on-target) effects, and not just on the causal pathway measured by the biomarker.

VALIDATING SURROGATE ENDPOINTS

To introduce his second main topic, the validation of surrogate endpoints, Dr. Fleming showed an attempt to use hematocrit levels as a surrogate endpoint for the risks of death and MI due to end-stage renal disease (Besarab et al., 1998). Standard-dose epogen, an erythropoiesis-stimulating agent (ESA), had been shown to partially normalize hematocrit levels, he said. Thus, a randomized trial was performed against more aggressive use of ESAs in patients with end-stage renal disease to see whether complete normalization of hematocrit would reduce their risk of death and MI. He noted that analyses of data in the standard-dose and in the high-dose arms indicated that every 10-point increase in hematocrit was associated with a 30 percent reduction in the relative risk of death. Since high-dose Epogen raised hematocrit by an average of nearly 10 points relative to standard-dose Epogen, one would expect that high-dose Epogen would reduce the death rate at least 25 percent as compared with the standard dose. However, nearly the opposite occurred: death rates of high-dose recipients increased 30 percent over those receiving the standard dose. "We now understand [these excess deaths] to be due to off-target effects, likely based in part on a thrombosis off-target mechanism," he said.

"A valid surrogate is one where the effect of the intervention on the surrogate is reliably telling us what the effect of the intervention is on the clinical endpoint," Dr. Fleming said. One setting in which a valid surrogate is being sought is type-2 diabetes. Hemoglobin A1c is a standard biomarker for blood sugar control. If an intervention reduced this biomarker by half a percent over a 6-month period, he asked, could we say that such an intervention would be effective in mitigating the long-term risk for microvascular and macrovascular complications in type 2 diabetes? Dr. Fleming noted that when this has been tried, the following major adverse effects have occurred:

- In the case of troglitazone, increased serious hepatic risks caused the drug to be taken off the market.
- Peroxisome proliferator-activated receptor agonists (muraglitazar and rosiglitazone) appeared to increase risk of death, stroke, and MI.
- The ACCORD trial found that an aggressive strategy to normalize hemoglobin A1c led to an increase in mortality.

Given these results, Dr. Fleming asked how one could elucidate effects on hemoglobin A1c that reliably predict clinical benefit? The Prentice criteria provide guidance, he said: first, the potential surrogate needs to be a correlate; second, the surrogate endpoint must fully capture the net effect of the intervention on all mechanisms that influence the clinical outcome. To determine circumstances in which hemoglobin A1c can act as a surrogate for type 2 diabetes, one can first examine the effect of the intervention on the clinical endpoint, such as the rate of cardiovascular death, stroke, and MI, he said. Then a statistical model, the Cox proportional hazards model, can be used to determine the proportion of net treatment effect explained by the surrogate endpoint. The key question is whether the effect of treatment on the clinical endpoint is fully captured by how treatment affects hemoglobin A1c levels.

Dr. Fleming emphasized that this question can best be answered by meta-analyses of many clinical studies in order to obtain sufficient evidence to determine whether or not the treatment's effect on the clinical endpoint is being fully captured by the effects on the biomarker. However, determination of the net effect of an intervention on a surrogate endpoint does not exclude the possibility that the intervention produces off-target effects on the clinical endpoint, he noted.

If an intervention for type 2 diabetes provided a 20 percent reduction in the rate of cardiovascular death, stroke, and MI, and that level effect exactly matched what is predicted based on its effect on hemoglobin A1c, it does not mean that is the only way that treatment influenced outcome, Dr. Fleming said. The intervention may have provided additional benefits via other causal pathways. Such off-target benefits could also be offset by adverse off-target effects that raise the risk of serious cardiovascular complications. Under these circumstances, the net effects on the surrogate and clinical endpoints would appear to be identical, but the biological reality would be very different. Echoing a point made by Victor De Gruttola in earlier discussion, he noted that "you can never discern whether the effect on the biomarker is capturing the totality of the effects [because] the effects are always going to be confounded … you are only able to assess the net effect."

In subsequent discussion, Richard Kuntz asked if these confounders could be reduced by intentionally studying the effects of different interventions on a biomarker. Dr. Fleming replied that if many different classes of interventions with diverse mechanisms produced the same effect on an endpoint of interest, that evidence would support the biomarker's validity. However, he said such evidence appeared to support lipid effects as a surrogate for CHD, until the effects of torcetrapib were understood. "Thank goodness we had clinical endpoint studies that reflected the totality of the effect," he said.

Such cases make clear that "biology does matter," Dr. Fleming said. "From the clinical perspective, it is key to have a comprehensive understanding of the causal pathways of the disease process, and of the off-target as well as the on-target effects of the intervention," he said. He noted that decades ago, it was generally thought that better statistics were needed to be able to validate surrogate endpoints. Today, he said, the evidence suggests that we need a richer clinical understanding of the disease processes and of the mechanisms of the intervention. Because it is "almost impossible" to obtain a complete biological understanding of a complex chronic disease, meta-analyses of clinical data provide the best route to such insights, he said.

One example of this is in the adjuvant colon cancer setting, when cancer is detected early enough to permit curative surgery. Among patients whose excised tissue contains positive nodes, about half will experience recurrence of cancer, due to microscopic undetected residual disease, and will die within 5 years, Dr. Fleming said. Therefore, patients with positive nodes receive chemotherapy to eradicate residual disease. He described a meta-analysis of 18 randomized clinical trials that was used to determine if delaying or reducing recurrence of disease is a valid surrogate for improving survival (Alonso and Molenberghs, 2008). The results suggested that to the extent that an intervention delayed or reduced recurrence, it had a proportionate effect on survival, he said. While this evidence supports the validity of the proposed surrogate endpoint, it also demonstrates that "surrogates tend to work best where you need them the least," Dr. Fleming said. In this case, the surrogate appears to be so close to the clinical endpoint as to provide modest advantage to its use.

In the HIV setting, Dr. Fleming said it is important to understand when a biomarker, such as viral load, may be used as a surrogate endpoint. For example, if viral load is lowered to undetectable levels in patients for 1 year, he said "it is probably going to be a pretty reliable surrogate for the ability to influence symptomatic AIDs-defining events and death if it is used in individuals with CD4 counts below 150." However, such a surrogate would have much greater clinical utility if it could be used to determine whether interventions to reduce viral load would

improve the long-term prognosis for newly infected people. Dr. Fleming said this is a much more complicated question.

A BIOMARKER HIERARCHY

Dr. Fleming has developed a four-level hierarchy for outcome measures, depending on the levels of evidence available:

- True clinical efficacy measures;
- Validated surrogate endpoints;
- Nonvalidated surrogate endpoints that are "reasonably likely to predict clinical benefit"; and
- Correlates that are solely a measure of biological activity (Fleming, 2005).

Dr. Fleming added that stroke might be a surrogate endpoint for overall survival in patients with atrial fibrillation, yet it also is a true clinical efficacy measure in that setting. He said that validated surrogate endpoints, the second level in the hierarchy, are relatively rare and include the earlier example of colon cancer recurrence and survival, as well as blood pressure, as a surrogate for clinical endpoints in antihypertensive interventions. Both of these surrogates were validated based on large amounts of data from multiple clinical trials, and they were validated for specific types of interventions used in those trials and for specific clinical endpoints, he said. In addition, the magnitude of each intervention's effect on the surrogate endpoint accurately predicted its effect on the clinical endpoint.

Most biomarkers occupy the two lowest levels of Dr. Fleming's hierarchy: those that are "reasonably likely to predict clinical benefit" and those that merely correlate with clinical benefit. Dr. Fleming emphasized that the definition "reasonably likely to predict clinical benefit" is an important distinction from a regulatory perspective because the accelerated approval process can be used in such settings. Biomarkers that attain the third level have the following attributes (beyond correlation with the clinical endpoint) according to Dr. Fleming:

- They accurately capture the treatment's effect on the predominant mechanism through which the disease process induces clinical risks.
- They are likely to capture large treatment effects on the clinical endpoint.
- They make predictions consistent with the net effect of an intervention on the clinical endpoint.

- They produce a target effect that is sufficiently strong and durable to enable them to predict meaningful benefit.

Biomarkers that correlate with clinical endpoints, but cannot predict them with reasonable likelihood, still may have many important uses, Dr. Fleming said. These include diagnosis of disease and assessing prognosis (as is the case for PSA); informing patient-specific therapeutic strategies (for example, adapting treatment for patients with pneumonia based on their body temperature, a biomarker for clinical benefit); serving as a primary endpoint for proof-of-concept or screening trials; and as an additional supportive measure of biological activity in phase III clinical trials. None of these uses is controversial, he pointed out; what is controversial, is using a biomarker as a surrogate endpoint.

Dr. Fleming also briefly discussed another purpose for biomarkers, patient enrichment, which is a strategy to identify and select patients who are likely to respond to a given intervention, such as a targeted therapeutic. Dr. Fleming characterized this use as very complicated, and noted that enrichment is typically used when the "key mechanisms of treatment effect on the causal factors of the disease process are specific to a targeted population." Examples of this include trastuzumab for patients whose breast cancers overexpress HER2, as well as cetuximab treatment for nonmutated k-ras tumors in colorectal cancer. In these types of situations, he noted, validation is particularly complex, as it is necessary to confirm that the "enriched" population defined using the biomarker responds differently than the general population; in addition, a robust assay must be able to define the target population.

John A. Wagner asked Dr. Fleming to comment on the variable utility of clinical endpoints, since some clinical endpoints suffer from many of the same measurement issues that biomarkers and surrogate endpoints do, as well as similar unintended consequences. For example, a PRO can measure how depressed patients feel, but it cannot capture the effect of a treatment on blood pressure or on suicide. Thus, he said, "I think it's important to take a bit of a look at the big picture for endpoints in general, and not just surrogate endpoints."

"My sense, and what the [IOM report] clearly points out, is the goal of clinical research is to enhance benefit to risk for the public," said Dr. Fleming. "That means basically improving how a patient feels, functions, and survives." Patients do not take therapy to alter their biomarkers; rather "they take therapy specifically to alter their clinical risk," he said.

However, he agreed with Dr. Wagner that some PROs may not capture important outcomes to patients, and said it is critically important to define endpoints that can reflect what matters to patients. Dr. Fleming reemphasized the need to select endpoints that are sensitive, clinically

relevant, and ideally, as comprehensive as possible. In life-threatening situations, Dr. Fleming said that survival often is the best endpoint to use, and there are many advantages to its use, since it is the easiest endpoint to validate and capture fully. In non-life-threatening situations, survival would not be the endpoint of choice, said Dr. Fleming. However, "in any setting, there could be clinically tangible effects that aren't captured by the primary endpoint," he added. "Therein lies the primary value of secondary endpoints … [of which] there should be a small number, because you otherwise run the risk of exploring the data and looking for those things that make you feel better about benefit to risk."

CONCLUDING THOUGHTS

Dr. Fleming addressed the consequences of relying on biomarkers as surrogate endpoints in the regulatory setting. He noted that natalizamab, a drug for multiple sclerosis, was given accelerated approval because it reduced the rate of relapse within a year; however, it was later associated with increased risk for progressive multifocal leukoencephalopathy (PML), a rare but serious brain infection. Two previously discussed treatments for type 2 diabetes—muraglitazar and rosiglitazone—were evaluated for full approvals on the basis of their effects on hemoglobin A1c, he added. These drugs have been associated with increased risk for death, stroke, and MI. These examples demonstrate that "when we're using surrogates, we're not only getting less reliable evidence about efficacy, we're also getting less reliable evidence about safety," Dr. Fleming said. Furthermore, because "everything is benefit-to-risk, the more limited information you have about the level of efficacy, the less resilient you are" when safety issues emerge. In the case of natalizamab, had trials been conducted that established beneficial effects in delaying clinical progression to walking with a cane or to being wheelchair bound, such effects on measures of irreversible morbidity would have provided much greater confidence in use of the agent even when evidence of PML emerged.

With regard to foods, Dr. Fleming noted that it is important to understand the net effect of an entire food on both biomarkers and clinical endpoints. Thus, he said, if the food in question has more than one experimental ingredient, it would be important to understand how each ingredient affects relevant biomarkers and clinical endpoints.

Addressing specific questions posed to him by the committee, Dr. Fleming revisited why it is important that each surrogate endpoint be evaluated on a case-by-case basis, and he discussed why a biomarker cannot be deemed a generic surrogate for a disease. "We have talked about several reasons why the generalizing of surrogates can be problematic," he said. He illustrated this with the following scenario: an intervention is

known to effectively reduce low-density lipoprotein cholesterol (LDL-C), and investigators want to use LDL lowering as a surrogate endpoint for death or MI. The new intervention has a similar effect on LDL, but it has far weaker effects on other positive mechanisms as compared with the original intervention. This could be because the original intervention not only reduced LDL, but also positively affected triglycerides and high-density lipoprotein cholesterol (HDL-C). The new intervention appears more beneficial than it is if investigators solely judge it by the effect on LDL, because biomarkers used as surrogate endpoints may not take into effect the multiple causal pathways involved in a disease process, he said.

Conversely, the new intervention could have unintended adverse effects such as increasing blood pressure through the angiotensin-renin system. Then, the effect on the lipid-based biomarker does not represent the totality of effects. "That is in fact what we saw when we looked at torcetrapib," he said. "Fortunately we recognized that torcetrapib and atorvastatin … [produced] an adverse [net] effect because we had clinical endpoint studies."

Dr. Fleming also reminded the audience that the magnitude and duration of the effect of the intervention matters. For example, an intervention that has a modest effect on LDL-C may not produce a clinical benefit. On the other hand, "we've also seen with some of the surrogates, that if the effect is particularly profound, more isn't always better," he said. Examples of this scenario include hematocrit normalization with ESAs, reducing hemoglobin A1c in type 2 diabetes, and large reductions in blood pressure (Staessen et al., 2003).

In response to the committee's question—how biomarker evaluation effects the public—Dr. Fleming replied that biomarkers are of great interest because they allow for timely assessment of interventions. However, he added, "it is critically important that [assessments] not just be timely, but reliable." The ultimate goal of these assessments is not to give the public more choices but rather more informed choices, he said. In that regard, he described the report as "very enlightened" in its discussion of the steps involved in biomarker evaluation: validation, qualification, and utilization. In undergoing such evaluation, he said it happens more frequently than one might expect that the effect of an intervention on a biomarker fails to accurately predict its effect on a clinical endpoint.

"It is not so much the things we don't know that get us into trouble; it's the things that we do know that aren't so," said Dr. Fleming. He added that for the public, the most problematic aspect of biomarker use results when biomarkers that are not truly validated give us the impression that we understand a treatment effect when we do not.

8

Key Themes, Challenges, and Opportunities

Over the course of the workshop, participants noted many potential uses of biomarkers, recognizing both their promise and limitations. Discussion of and reaction to the committee's report coalesced around a few important topics, including

- The structure and usefulness of the biomarker evaluation framework;
- The value of biomarkers and surrogate endpoints;
- Considerations in the food and nutrition settings; and
- Communication of information to consumers and medical professionals.

BIOMARKER EVALUATION FRAMEWORK

Many speakers noted the importance of a biomarker evaluation framework across product areas. For example, Andrew Shao said that an absence of an accepted framework has limited the amount of research conducted on the role of diet and nutrition in health promotion and disease prevention: "We anticipate that a formal biomarker evaluation process will add clarity to product development, as companies that choose to invest in research will have a better understanding a priori that the research will have broader acceptability and applicability to public health recommendations." Stephen Williams said that the pharmaceutical industry wants a consistent, transparent biomarker evaluation framework that

would enable drug companies and regulators to decide what level of evidence may be required for a particular biomarker application, and acknowledged some aspects of the committee's recommended framework were steps in the right direction. Jack Zakowski agreed with the committee's biomarker evaluation framework, especially the focus on the interdependence of the three steps of the framework, and Guy Johnson noted that the framework was comprehensive. While there was general agreement on the need for a biomarker evaluation framework, several speakers expressed differing opinions on specific aspects of the framework and its implications.

James Mayne and Dr. Williams were concerned that the report did not specify criteria that should be applied to biomarkers at each step of the evaluation framework. "I eagerly tore through the document … looking for the actual elements of the decision framework, by what criteria would decisions be made in the regulatory space," said Dr. Mayne. "That was not provided, at least not in the detail I was looking for." John R. Ball noted that the committee didn't view their recommendations as the last word in biomarker evaluation but as a fulfillment of their charge to develop a framework for biomarker evaluation across the Food and Drug Administration (FDA) regulatory spectrum. He added that although the Center for Food Safety and Applied Nutrition's (CFSAN's) work would be simplified if the Institute of Medicine (IOM) committee had developed a five-item checklist of criteria that every biomarker used in a health claim had to fulfill, the committee found this notion unrealistic. Given limited understanding of chronic disease and the biological significance of existing biomarkers, the committee concluded that evaluation must be conducted by expert panels on a case-by-case basis, Dr. Ball said. For further discussion of the committee's evaluation framework and related recommendations, see Chapter 3 of its report (IOM, 2010).

The effect of the biomarker evaluation framework on innovation generated additional discussion. By employing a rigorous biomarker evaluation framework, there were concerns that this may unintentionally discourage research in the area of biomarkers. However, many speakers noted that the biomarker evaluation framework will not have a chilling effect on biomarker research and development. Thomas Fleming suggested that a lack of clarity, both from the regulatory and scientific perspective, on biomarker evaluation standards is much more likely to inhibit innovation. Implementation of Recommendation 3 of the report could help FDA to bring scientific and regulatory clarity to biomarker evaluation across the FDA's centers and regulated product categories. David DeMets added that the inappropriate use of a biomarker would have negative effects on innovation. Dr. Mayne said that the report clearly stated that the framework is not intended for biomarkers used in the discovery space: "I think

the committee made it clear that those areas are omitted so as not to constrain innovation and not to preconfer how or which biomarkers might be advanced." See also pages 100–102 and 116–121 of the committee's report (IOM, 2010).

One of the aspects of the biomarker evaluation framework that received substantial discussion was the third step of the biomarker evaluation framework, utilization. Several speakers questioned the placement of the utilization step after analytical validation and qualification. For example, Dr. Johnson said that for biomarkers used to support health claims, the context of use is very specifically defined in terms of the food vehicle, its target audience, and the setting in which it is consumed. It would therefore be difficult to separate biomarker utilization as applied to food from qualification, he said. Dr. Williams asserted that the specific context for biomarker use must be defined at the outset of evaluation. "If you have not defined [the context of use] then you don't know what the value of success would be, or the value of the truth, if the biomarker actually does what you think it does. And you don't know what the consequence of error would be … if it fails to do so," Dr. Williams said. These determinations drive tolerance to risk: "If the truth is extremely valuable and failure is inconsequential, then your tolerance for variation of all kinds is pretty good, and you don't require much evidence," he said. "But if there is not much value to the truth and there are terrible consequences to errors, then you are going to require a lot of information about precision of that biomarker and lot of evidence that actually it is going to do what you say it is." Therefore, he said "if you haven't defined the [biomarker's] purpose first, then I don't think you can define how good is good enough, and that's what validation and qualification are about."

Maria Lopes-Virella said that the committee recognized that there must be a rationale for embarking on the evaluation of a biomarker, which can, in some sense, be seen as utilization, which is the reason the framework was depicted as circular. In organizing the evaluation framework, the committee placed analytical validation, qualification, and utilization in order of the decisions that would be made, understanding that the biomarker evaluation process would be initiated on the basis of an initial motivation or context of use. The committee reasoned that only after passing the first two thresholds—validation and qualification—should decisions be made about whether or not to use a biomarker. However, if a biomarker cannot be reproducibly measured, or is otherwise analytically invalid, there is no point in evaluating it further, said Dr. Ball. He added that the evaluation framework is not a series of disjointed steps, but rather an integrated process. Further explanation of the committee's rationale for the order of evaluation framework steps can be found on pages 119–120 of its report (IOM, 2010).

Dr. Ball said that the Center for Drug Evaluation and Research's (CDER's) concept of qualification differs from the committee's because it includes consideration of a biomarker's context of use. "I think conceptually, because of the way CDER started with an integrated kind of approach [to biomarker evaluation], both industry and FDA conceptualize the process as integrated," he said. Alternatively, the committee conceptualizes the evaluation process as stepwise but always informed by context.

The order of the three steps also has a functional rationale, Allan Jaffe noted: existing analytical validation data "can be put together in a fairly facile way" before determining whether the biomarker can be used in a particular context. Dr. Jaffe said that defining a narrow context of use for a biomarker from the outset may limit its potential applications. Instead, "we start to investigate biomarkers, [conducting] the analytic validation first. We then look … at disease entities that have large numbers of individuals [who] we can study to develop some sort of relationship to outcomes, and then we'll look at other contexts of use."

THE VALUE OF BIOMARKERS AND SURROGATE ENDPOINTS

Throughout the discussion forum, speakers provided their perspectives on the value of biomarkers, especially those used as surrogate endpoints. Kathleen Ellwood noted that reliance on long-term clinical trials is not always feasible, and that the FDA requested the IOM to undertake this study to address the absence of an agreed-upon, systematic, and transparent process for qualifying surrogate endpoints. Several speakers from the food and nutrition industries said that they rely on biomarkers and surrogate endpoints to conduct trials on nutritional interventions, because it is too costly or logistically challenging to conduct trials with clinical endpoints. These speakers also highlighted the importance of observational studies of the impacts of foods on clinical outcomes in the bodies of evidence supporting claims on foods. In the drug development setting, Dr. Williams noted that surrogate endpoint or efficient clinical endpoint availability was associated with more therapies, as compared to diseases that lacked surrogate endpoints, according to a recent review.

Other stakeholders expressed concerns about the use of surrogate endpoints. Michael Lauer disagreed with the committee's statement that "biomarkers can enable faster, more efficient clinical trials for life-saving and health promoting interventions," and noted that perhaps the opposite of this statement is true. Dr. Lauer said that it is impossible to avoid clinical endpoint clinical trials to assess the effect of interventions. Dr. Fleming noted that validated surrogate endpoints are extremely rare; however, he added that there are many other uses of biomarkers that are critically important, including the diagnosis and prognosis of disease,

informing patient-specific therapeutic strategies, primary endpoints in proof-of-concept studies or screening trials, patient enrichment, and as additional measures of biological activity in phase III clinical trials.

CONSIDERATIONS FOR FOOD AND NUTRITION APPLICATIONS

Many speakers emphasized the importance of a biomarker evaluation framework. However, there was some discussion among speakers that the committee's framework and associated recommendations would be difficult to implement in a food or nutrition setting. Several stakeholders said that foods are different from other FDA-regulated product areas, and suggested that biomarker evaluations in foods require special consideration.

Dr. Shao said that the overall context of biomarker use is different in foods than in drugs. Foods are presumed to be safe, and while that doesn't mean they're risk free, he said their risk paradigm is very different from that of a cancer drug. Foods with health claims promote health and are not urgent interventions, he added. The health effects of foods are modest as compared with drugs, and they are spread over a heterogeneous population, as compared with patients prescribed a particular drug, Dr. Shao said. As noted in her earlier presentation, Roberta Ness said that the committee's rationale for recommending that the same degree of scientific rigor be applied across all regulatory settings—including food—resulted from an understanding that foods may not be implicitly free from harm. Dr. Ball added that the committee recognized that drugs are ingested by a small proportion of the population, and their use is guided by physicians, whereas foods are ingested by everyone, largely without guidance, and individuals may be less able to interpret both the risks and benefits associated with consuming a particular food. To this point, Dr. Johnson suggested that the public may indeed be exposed to risk through the addition of nutrients such as vitamin D to foods, but that eating larger amounts of individual, unfortified foods, such as strawberries, poses little risk to public health.

Douglas Balentine said that the process by which food and nutrition companies identify promising biomarkers is fundamentally different from the way the pharmaceutical industry develops biomarkers. Many food biomarker leads come from epidemiological or observational studies that examine "intake markers" such as beta-carotene or the consumption of certain foods. For example, results of a study of fiber intake and cholesterol levels might suggest that eating whole grains is associated with a reduction in risk of coronary vascular disease and death, and blood samples from this study also suggest that individuals with lower risk have lower values of the biomarker, low-density lipoprotein cholesterol

(LDL-C). The food industry would like to take this information and use it to pursue studies on foods, or on food fiber in purified form, to look at the effect of fiber on LDL-C, he said. To do so, they need to know if LDL-C is a valid biomarker, and if not, what kind of evidence is needed to support its use as a biomarker to examine foods that claim to offer health benefits. Observational studies drive the identification of clinical endpoints, Dr. Balentine said. Food companies are unlikely to develop a biomarker because to do so would involve proving that the link to the clinical endpoint is valid, he said. If that is the case, there is little need for a biomarker. "We need rigorous science and … appropriate trials that you can do within foods and complex foods," he stated.

"The food industry does all kinds of randomized studies," Dr. Johnson said, noting that such studies are required by the FDA to support health claims. However, randomized studies of food "cannot be based on a hard clinical endpoint. You cannot randomize people to Cheerios or cornflakes and see who gets heart disease 20 years down the pike," he said. Thus, he had hoped that the committee would recommend "ways to use surrogate endpoints to inform shorter-term clinical trials that the food industry could do with an endpoint that made sense."

Victor De Gruttola responded that "if you can only do the randomized study on the biomarker, but not the clinical endpoint, then you are in the world of observational studies." Such studies show a causal effect on the biomarker, but to conclude that there is a causal impact on the clinical endpoint requires a randomized controlled trial, he said—unless the biomarker has been validated as a surrogate endpoint. If evidence from a controlled trial of foods shows an effect on a biomarker, and epidemiologic evidence suggests that the effect on the biomarker may be correlated with the effect on the clinical endpoint, this information leads to a hypothesis, Dr. Fleming added. "That is where we were with beta-carotene," he said: the hypothesis was that increasing beta-carotene levels should reduce disease incidence (see also section on beta-carotene qualification in IOM, 2010). This hypothesis turned out to be incorrect, and that determination could only have been made through a randomized controlled trial that studied clinical outcomes.

While acknowledging the difficulty of conducting large-scale clinical endpoint trials with foods, Dr. Fleming said that foods can be marketed without such claims, unlike drugs. If you want to make that claim for any product, then reliable evidence for that claim can only be obtained through large-scale clinical or validated surrogate endpoint trials, he said. "There are many different ways foods could be beneficial for public health," Dr. Fleming said. However, to claim that a food prevents cancer, for example, requires the same kind of reliable evidence as would a drug making the same claim, he asserted.

Most health claims involve cardiovascular disease because FDA considers LDL-C a valid biomarker, Dr. Johnson stated. If there were other biomarkers that were similarly validated, the food industry could use them to make a positive impact on public health, he said. Dr. Lopes-Virella said that it would be more scientifically valid to make a claim stating effect on cholesterol rather than clinical outcome if the data supporting the claim are based on biomarker endpoints. If the claim is not based on clinical endpoints, "don't say that it reduces heart disease," she insisted.

Some speakers noted that there is a lack of incentives for the food and nutrition industries to conduct rigorous studies to support the biomarker evaluation framework. Furthermore, speakers noted that nutrition-based clinical trials are complicated by a number of circumstances, including individuals in the control arms who inadvertently consume the intervention, the complex and highly variable food matrices that can modify nutrient bioavailability, and overall dietary patterns. Dr. Shao said these factors also make running and interpreting trials for foods and supplements extremely difficult and complex. Dr. Ball responded that the committee recognized that the food industry lacks the profit margins, financial incentives, and logistics to carry out studies of biomarkers similar to those conducted by the pharmaceutical industry. Because of these disincentives, the committee felt strongly about recommending that the Department of Health and Human Services (HHS) should facilitate a coordinated, department-wide effort to encourage the sharing of data about biomarkers, he said (see also Chapter 5 of the committee's report [IOM, 2010]). Dr. Jaffe noted that he understood that there is a lack of clarity of how to operationalize this recommendation, especially in the food settings. However, he added that having a unified framework could help the food and supplement industries by leveraging the biomarker evaluation experiences of the drug and device industries.

IMPROVING COMMUNICATION AND UNDERSTANDING

Another theme addressed in the report and raised several times in discussion is the challenge of communicating information about biomarkers—and science in general—to the public. The validity of claims that are made on the basis of biomarkers depends in part upon how they are understood by patients, consumers, and health care workers, Dr. Ball said. However, lack of numeracy among medical professionals, as well as among the general public, presents a major obstacle to consumer understanding and interpretation of claims based on biomarkers, he noted. It is not enough to simply provide numbers, he said; we need to be aware of how people understand the information they receive and how they act in response.

Dr. Jaffe added that the place to start making efforts to communicate

such information better at all levels is at medical schools, during internships and residencies, where physicians can be taught to think probabilistically. Dr. Zakowski observed that nearly everyone is innumerate in some aspect of their decision making, but that media interest in this report—and in scientific developments in general—represents an opportunity to teach the importance of numeracy.

Dr. Khoo added that scientists need to better communicate among themselves regarding the implications of their work. "I think we also need to educate the food science discipline [regarding biomarkers] … because where the food industry is going to have to digest this information is, in many ways, in product development," she said. For example, there are multiple ways food scientists might approach a surrogate endpoint such as blood pressure. Biologists and medical scientists need to agree among themselves what a biomarker means, rather than present conflicting or confusing information to the public or the media, she said.

"I certainly want to support the notion that communication as perhaps the ultimate end product of this work is extremely important," Dr. Mayne said. He urged that all stakeholders be included in educational efforts involving biomarkers, because "these concepts are not easy for anyone to get their brain around." He further suggested that a "layered approach" to communicating information about biomarkers could help people understand how it applies to their own choices, and how they can use it to "frame their expectations, whether they are taking a box of cereal off the shelf or whether they are trying to advance a new diagnostic."

Dr. Ball said he was struck by the interest of both the lay and trade press in the report, which he thought reflected the importance of healthy eating to the public. That issue is unlikely to fade, he added, and it ought to be addressed by both regulatory and industry sectors.

Dr. Johnson noted that one of the fundamental goals of the Nutrition Labeling and Education Act (NLEA) of 1990 was to provide people with the information they need to select a healthful diet. The food industry has "a tremendous potential to communicate positive, useful, helpful information," he said. In addition, he said that food companies understand that claims not based on sound science will backfire.

Melissa Musiker said the Grocery Manufacturers Association (GMA) was equally surprised by the degree of media interest in the report. She emphasized that consumers fail to distinguish among the different types of health claims. The GMA puts considerable effort into determining how to make a structure–function claim that is truthful and communicates appropriate information to consumers, she said.

Dr. Balentine noted that Unilever would welcome improvements in numeracy that would allow consumers to better distinguish among products with health claims. For example, their products contain what they

believe to be an effective dose of plant sterols, based on the results of clinical trials, while other companies put far smaller amounts of sterols in comparable items. Unfortunately, he said, the claims that Unilever is able to make today do not differentiate their product as containing a higher level of this more expensive ingredient.

Communication was one of the main topics the committee discussed, according to Dr. Lopes-Virella, and committee members were concerned that their message be clear and not frightening. Most importantly, she said people need to understand that biomarker-based information, like nearly all medical information, conveys probabilities. "Everyone needs to understand there are no absolutes in medicine," she emphasized.

9

Importance of the Biomarker Discussion Forum

At the conclusion of the discussion forum, John R. Ball reaffirmed the committee's goals in convening this event. He noted that the discussion forum provided an opportunity for stakeholders to learn more about the report and enabled the committee to further flesh out the rationale behind their recommendations. Importantly, the committee was also able to hear from the stakeholders affected by the report recommendations. Dr. Ball noted that these discussions were extremely valuable, and he hoped that this setting allowed all stakeholders—including industry representatives and regulators—to have a dialogue about the important issues in biomarker evaluation.

References

Alonso, A., and G. Molenberghs. 2008. Evaluating time to cancer recurrence as a surrogate marker for survival from an information theory perspective. *Statistical Methods in Medical Research* 17(5):497–504.

Antman, E. M., M. Hand, P. W. Armstrong, E. R. Bates, L. A. Green, L. K. Halasyamani, J. S. Hochman, H. M. Krumholz, G. A. Lamas, C. J. Mullany, D. L. Pearle, M. A. Sloan, S. C. Smith, Jr., D. T. Anbe, F. G. Kushner, J. P. Ornato, A. K. Jacobs, C. D. Adams, J. L. Anderson, C. E. Buller, M. A. Creager, S. M. Ettinger, J. L. Halperin, S. A. Hunt, B. W. Lytle, R. Nishimura, R. L. Page, B. Riegel, L. G. Tarkington, and C. W. Yancy. 2008. 2007 focused update of the ACC/AHA 2004 guidelines for the management of patients with ST-elevation myocardial infarction: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Journal of the American College of Cardiology* 51(2):210–247.

Baigent, C., A. Keech, P. M. Kearney, L. Blackwell, G. Buck, C. Pollicino, A. Kirby, T. Sourjina, R. Peto, R. Collins, and R. Simes. 2005. Efficacy and safety of cholesterol-lowering treatment: Prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins. *Lancet* 366(9493):1267–1278.

Bardy, G. H., K. L. Lee, D. B. Mark, J. E. Poole, D. L. Packer, R. Boineau, M. Domanski, C. Troutman, J. Anderson, G. Johnson, S. E. McNulty, N. Clapp-Channing, L. D. Davidson-Ray, E. S. Fraulo, D. P. Fishbein, R. M. Luceri, and J. H. Ip. 2005. Amiodarone or an implantable cardioverter-defibrillator for congestive heart failure. *New England Journal of Medicine* 352(3):225–237.

Barter, P. J., M. Caulfield, M. Eriksson, S. M. Grundy, J. J. 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.

Besarab, A., W. K. Bolton, J. K. Browne, J. C. Egrie, A. R. Nissenson, D. M. Okamoto, S. J. Schwab, and D. A. Goodkin. 1998. The effects of normal as compared with low hematocrit values in patients with cardiac disease who are receiving hemodialysis and epoetin. *New England Journal of Medicine* 339(9):584–590.

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.

Califf, R. M. 2006. Clinical trials bureaucracy: Unintended consequences of well-intentioned policy. *Clinical Trials* 3(6):496–502.

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. *New England Journal of Medicine* 321(6):406–412.

Castelli, W. P. 1988. Cholesterol and lipids in the risk of coronary artery disease—The Framingham Heart Study. *Canadian Journal of Cardiology* 4(Suppl A):5A–10A.

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).

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.

de Lorgeril, M., P. Salen, J. L. Martin, I. Monjaud, J. Delaye, and N. Mamelle. 1999. Mediterranean diet, traditional risk factors, and the rate of cardiovascular complications after myocardial infarction: Final report of the Lyon Diet Heart Study. *Circulation* 99(6):779–785.

Dieterle, F., F. Sistare, F. Goodsaid, M. Papaluca, J. S. Ozer, C. P. Webb, W. Baer, A. Senagore, M. J. Schipper, J. Vonderscher, S. Sultana, D. L. Gerhold, J. A. Phillips, G. Maurer, K. Carl, D. Laurie, E. Harpur, M. Sonee, D. Ennulat, D. Holder, D. Andrews-Cleavenger, Y. Z. Gu, K. L. Thompson, P. L. Goering, J. M. Vidal, E. Abadie, R. Maciulaitis, D. Jacobson-Kram, A. F. Defelice, E. A. Hausner, M. Blank, A. Thompson, P. Harlow, D. Throckmorton, S. Xiao, N. Xu, W. Taylor, S. Vamvakas, B. Flamion, B. S. Lima, P. Kasper, M. Pasanen, K. Prasad, S. Troth, D. Bounous, D. Robinson-Gravatt, G. Betton, M. A. Davis, J. Akunda, J. E. McDuffie, L. Suter, L. Obert, M. Guffroy, M. Pinches, S. Jayadev, E. A. Blomme, S. A. Beushausen, V. G. Barlow, N. Collins, J. Waring, D. Honor, S. Snook, J. Lee, P. Rossi, E. Walker, and W. Mattes. 2010. Renal biomarker qualification submission: A dialog between the FDA-EMEA and Predictive Safety Testing Consortium. *Nature Biotechnology* 28(5):455–462.

Dluhy, R. G., and G. T. McMahon. 2008. Intensive glycemic control in the ACCORD and ADVANCE trials. *New England Journal of Medicine* 358(24):2630–2633.

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 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.

FDA (Food and Drug Administration). 2009. *Summary of qualified health claims subject to enforcement discretion*. http://www.fda.gov/Food/LabelingNutrition/LabelClaims/QualifiedHealthClaims/ucm073992.htm#gtea (accessed October 12, 2010).

Fleming, T. R. 2005. Surrogate endpoints and FDA's accelerated approval process. *Health Affairs (Millwood)* 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.

Gallin, J. I., H. L. Malech, D. A. Melnik, R. S. Weening, D. Roos, H. Wendt, R. C. Woodman, J. T. Curnutte, R. A. Seger, T. Muhlebach, J. P. Bourquin, R. A. Axtell, L. A. Boxer, C. L. Pierson, D. C. Anderson, W. T. Shearer, H. M. Rosenblatt, J. C. Torkildson, R. L. Baehner, T. D. Coates, K. Belani, P. G. Quie, N. H. Valerius, C. Koch, F. K. Pedersen, H. R. Hill, A. Shigeoka, S. H. Fischer, H. Rosen, H. D. Ochs, T. R. Fleming, C. Cunningham Rundles, A. Hyatt, H. Evrard, C. Nathan, C. Farber, R. A. B. Ezekowitz, A. E. Izu, S. M. Kramer, and H. S. Jaffe. 1991. A controlled trial of interferon gamma to prevent infection in chronic granulomatous disease. *New England Journal of Medicine* 324(8):509–516.

GAO (Government Accountability Office). 2000. *Food safety: Improvements needed in overseeing the safety of dietary supplements and "functional foods."* Washington, DC: GAO.

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

Gordon, D. J. 1994. Cholesterol lowering and total mortality. In *Contemporary issues in cholestorol lowering: Clinical and population aspects*, edited by B. M. Rifkind. New York: Marcel Dekker.

Gordon, T., W. P. Castelli, M. C. Hjortland, W. B. Kannel, and T. R. Dawber. 1977. High-density lipoprotein as a protective factor against coronary heart disease: The Framingham Study. *American Journal of Medicine* 62(5):707–714.

HHS (Department of Health and Human Services). 2005. *Dietary Guidelines for Americans*. http://www.health.gov/dietaryguidelines/dga2005/document/pdf/DGA2005.pdf (accessed December 15, 2010).

Hochman, J. S., G. A. Lamas, C. E. Buller, V. Dzavik, H. R. Reynolds, S. J. Abramsky, S. Forman, W. Ruzyllo, A. P. Maggioni, H. White, Z. Sadowski, A. C. Carvalho, J. M. Rankin, J. P. Renkin, P. G. Steg, A. M. Mascette, G. Sopko, M. E. Pfisterer, J. Leor, V. Fridrich, D. B. Mark, and G. L. Knatterud. 2006. Coronary intervention for persistent occlusion after myocardial infarction. *New England Journal of Medicine* 355(23):2395–2407.

IOM (Institute of Medicine). 2006. *Dietary reference intakes: The essential guide to nutrient requirements*. 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 and protecting the health of the public*. Washington, DC: The National Academies Press.

IOM. 2010. *Evaluation of biomarkers and surrogate endpoints in chronic disease.* Washington, DC: The National Academies Press.

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.

Lewis, M. M. 2003. *Moneyball: The art of winning an unfair game*. New York: W.W. Norton.

Mauri, L., E. J. Orav, and R. E. Kuntz. 2005. Late loss in lumen diameter and binary restenosis for drug-eluting stent comparison. *Circulation* 111(25):3435–3442.

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 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.

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

Robinson, J. G., and N. J. Stone. 2006. Identifying patients for aggressive cholesterol lowering: The risk curve concept. *American Journal of Cardiology* 98(10):1405–1408.

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.

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

Seshadri, S., A. L. Fitzpatrick, M. A. Ikram, A. L. DeStefano, V. Gudnason, M. Boada, J. C. Bis, A. V. Smith, M. M. Carassquillo, J. C. Lambert, D. Harold, E. M. Schrijvers, R. Ramirez-Lorca, S. Debette, W. T. Longstreth, Jr., A. C. Janssens, V. S. Pankratz, J. F. Dartigues, P. Hollingworth, T. Aspelund, I. Hernandez, A. Beiser, L. H. Kuller, P. J. Koudstaal, D. W. Dickson, C. Tzourio, R. Abraham, C. Antunez, Y. Du, J. I. Rotter, Y. S. Aulchenko, T. B. Harris, R. C. Petersen, C. Berr, M. J. Owen, J. Lopez-Arrieta, B. N. Varadarajan, J. T. Becker, F. Rivadeneira, M. A. Nalls, N. R. Graff-Radford, D. Campion, S. Auerbach, K. Rice, A. Hofman, P. V. Jonsson, H. Schmidt, M. Lathrop, T. H. Mosley, R. Au, B. M. Psaty, A. G. Uitterlinden, L. A. Farrer, T. Lumley, A. Ruiz, J. Williams, P. Amouyel, S. G. Younkin, P. A. Wolf, L. J. Launer, O. L. Lopez, C. M. van Duijn, and M. M. Breteler. 2010. Genome-wide analysis of genetic loci associated with Alzheimer disease. *Journal of the American Medical Association* 303(18):1832–1840.

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.

Torpy, J. M., C. Lynm, and R. M. Glass. 2007. JAMA patient page. Implantable cardioverter-defibrillators. *Journal of the American Medical Association* 297(17):1946.

Trumbo, P., and K. Ellwood. 2009. *Developing a framework for biomarker qualification for chronic disease*. Paper presented at Committee on qualification of biomarkers and surrogate endpoints in chronic disease, Washington, DC.

Varmus, H. 2010. Ten years on—The human genome and medicine. *New England Journal of Medicine* 362(21):2028–2029.

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

Wang, T. J., P. Gona, M. G. Larson, G. H. Tofler, D. Levy, C. Newton-Cheh, P. F. Jacques, N. Rifai, J. Selhub, S. J. Robins, E. J. Benjamin, R. B. D'Agostino, and R. S. Vasan. 2006. Multiple biomarkers for the prediction of first major cardiovascular events and death. *New England Journal of Medicine* 355(25):2631–2639.

Yetley, E. A., C. M. Pfeiffer, R. L. Schleicher, K. W. Phinney, D. A. Lacher, S. Christakos, J. H. Eckfeldt, J. C. Fleet, G. Howard, A. N. Hoofnagle, S. L. Hui, G. L. Lensmeyer, J. Massaro, M. Peacock, B. Rosner, D. Wiebe, R. L. Bailey, P. M. Coates, A. C. Looker, C. Sempos, C. L. Johnson, and M. F. Picciano. 2010. NHANES monitoring of serum 25-hydroxyvitamin D: A roundtable summary. *Journal of Nutrition* 140(11):2030S–2045S.

Acronyms

ACCORD	Action to Control Cardiovascular Risk in Diabetes
ACE inhibitor	angiotensin-converting enzyme inhibitor
ADA	American Dietetic Association
APOE	apolipoprotein E
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
CEA	carcinoembryonic antigen
CFSAN	Center for Food Safety and Applied Nutrition
CHD	coronary heart disease
CLIA	Clinical Laboratory Improvement Amendments
CMS	Centers for Medicare & Medicaid Services
COAG	Clarification of Optimal Coagulation through Genetics
COPD	chronic obstructive pulmonary disease
CRN	Council for Responsible Nutrition
CRP	C-reactive protein
CSPI	Center for Science in the Public Interest
CVD	cardiovascular disease

DDT	drug development tools
DHA	docosahexaenoic acid
EMEA	European Medicines Agency
ESA	erythropoiesis-stimulating agent
FAA	Federal Aviation Administration
FDA	Food and Drug Administration
FDAAA	Food and Drug Administration Amendments Act
GAO	Government Accountability Office
GI	gastrointestinal
GMA	Grocery Manufacturers Association
HDL	high-density lipoprotein
HDL-C	high-density lipoprotein cholesterol
HER2	human epidermal growth factor receptor 2
HHS	Department of Health and Human Services
HIV	human immunodeficiency virus
HRT	hormone replacement therapy
IOM	Institute of Medicine
IVD	*in vitro* diagnostic
LDL	low-density lipoprotein
LDL-C	low-density lipoprotein cholesterol
MI	myocardial infarction
MRI	magnetic resonance imaging
NCI	National Cancer Institute
NHANES	National Health and Nutrition Examination Survey
NHLBI	National Heart, Lung, and Blood Institute
NIH	National Institutes of Health
NLEA	Nutrition, Labeling, and Education Act
ODS	Office of Dietary Supplements
PCI	percutaneous coronary intervention
PET	positron emission tomography
PML	progressive multifocal leukoencephalopathy
PRO	patient-reported outcomes
PSA	prostate-specific antigen

PSTC	Predictive Safety Testing Consortium
RT-PCR	reverse-transcriptase polymerase chain reaction
SABRe	Systems Approach to Biomarker Research in Cardiovascular Disease
SPECT	single photon emission computed tomography
TLR	target lesion revascularization
TSH	thyroid-stimulating hormone
USDA	United States Department of Agriculture
VRE	vancomycin-resistant enterococci

Glossary

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 (FDA) using earlier clinical trial results than would be required for regular approvals; postmarket surveillance and studies generally required

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 high blood pressure; prevents formation of a protein that causes constriction of blood vessels, thus lowering blood pressure

Apolipoprotein—a protein component of lipoprotein complexes

Assay—a biochemical or other measurement developed to quantitate a biomarker

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

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)

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 FDA, 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"

Calcium channel blocker—drug used to 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

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

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

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 [or consumer] 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)

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)

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

Correlation—a statistical association between two variables that does not imply a cause-and-effect relationship

C-reactive protein (CRP)—an acute-phase, nonspecific, systemic biomarker 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

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)

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 FDA review or authorization before use

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

Efficacy—ability to produce a desired effect

Epidemiologic studies—studies of the health of various human populations

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)

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

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

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)

In vitro diagnostic—a test that can detect disease, infection, or other health conditions

In vivo—in the living body of a plant or animal

Intervention—any drug, device, biologic, behavioral modification, nutritional modification, lifestyle modification, or other treatment intended to improve health

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

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)

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

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

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*

Pathophysiology—processes leading to the incidence or progression of disease or other health-related condition; alteration in function as distinguished from structural defects

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; FDA generally requires new drugs to be tested in phase III trials before they can be put on the market (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 FDA for already approved drugs or devices to review potential risks

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

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

Risk–benefit analysis—the comparison of the risk of a situation to its benefit

Risk biomarker—a biomarker that indicates a risk factor for a disease

Risk factors—variables that predict outcomes and are composed of biomarkers and social and environmental factors

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

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

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 (Biomarkers Definitions Working Group, 2001)

Surveillance—population-level monitoring for early detection and treatment of advancing disease or complications

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)

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

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

Appendix A

Discussion Forum Agenda

Evaluation of Biomarkers and Surrogate Endpoints in Chronic Disease

June 21–22, 2010
Washington Plaza Hotel
10 Thomas Circle, NW
Washington, DC

Day 1, June 21

8:30 am	**Session 1: IOM Report Recommendations**
8:30 am	*Committee member presentations* John R. Ball, Welcome, the Committee's Work, and Workshop Objective
8:50 am	John A. Wagner, Recommendations 1–2, the Evaluation Framework
9:10 am	Ronald Krauss and Maria Lopes-Virella, LDL and HDL Case Study
9:30 am	Roberta Ness, Recommendations 3–6
9:50 am	*Panel discussion*
10:30 am	**Session 2: FDA and NIH Perspectives**
10:30 am	Kathleen Ellwood and Paula Trumbo, CFSAN: Current biomarker processes at FDA and basis for IOM study
10:45 am	Michael Lauer, Division of Cardiovascular Sciences, NHLBI, NIH
11:00 am	Paul Coates, NIH Office of Dietary Supplements
11:15 am	BREAK
11:30 am	Marc Walton, CDER: Biomarker Qualification in CDER
11:45 am	Robert Becker, CDRH: Biomarkers for Devices
12:00 pm	LUNCH
12:15 pm	*Session 2 Panel discussion* Moderated by Roberta Ness

1:00 pm	**Session 3: Stakeholder Reaction and Discussion** *5-minute presentations from stakeholders*
1:05 pm	Douglas Balentine, Unilever, conventional food
1:10 pm	Guy Johnson, Johnson Nutrition Solutions, conventional food
1:15 pm	Andrew Shao, Council for Responsible Nutrition, supplements industry
1:20 pm	Mary Hager, American Dietetic Association, food and nutrition professional organization
1:25 pm	Ilene Heller, Center for Science in the Public Interest, food consumer advocacy organization
1:30 pm	Stephen Williams, Somalogic, pharmaceutical industry
1:35 pm	Jack Zakowski, Beckman Coulter, diagnostics industry
1:40 pm	Richard Kuntz, Medtronic, device industry
1:45 pm	*Panel discussion* Moderated by John A. Wagner
2:30 pm	BREAK
2:45 pm	Invited Overview Presentation, Thomas Fleming
3:30 pm	**Session 4: Implications and Discussion** *Panel discussion on report recommendations moderated by John R. Ball*
5:00 pm	Wrap-up John R. Ball and the Committee
5:00 pm	ADJOURN Day 1

Day 2, June 22

9:00 am	Welcome and Recap of Day 1 John R. Ball, committee chair
9:30 am	**Session 5: Session for Stakeholder Comments** *5-minute reactions to report recommendations*
9:45 am	James Mayne, Pfizer
9:50 am	Melissa Musiker, Grocery Manufacturers Association
9:55 am	Federico Goodsaid, Food and Drug Administration
11:45 am	Wrap-up John R. Ball, committee chair
12:00 pm	ADJOURN MEETING

Appendix B

Summary from the Committee's Report *Evaluation of Biomarkers and Surrogate Endpoints in Chronic Disease*

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 nonuniform 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 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 a low-density lipoprotein cholesterol (LDL-C) level (more accurately, non-high-density lipoprotein cholesterol [HDL-C]), 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

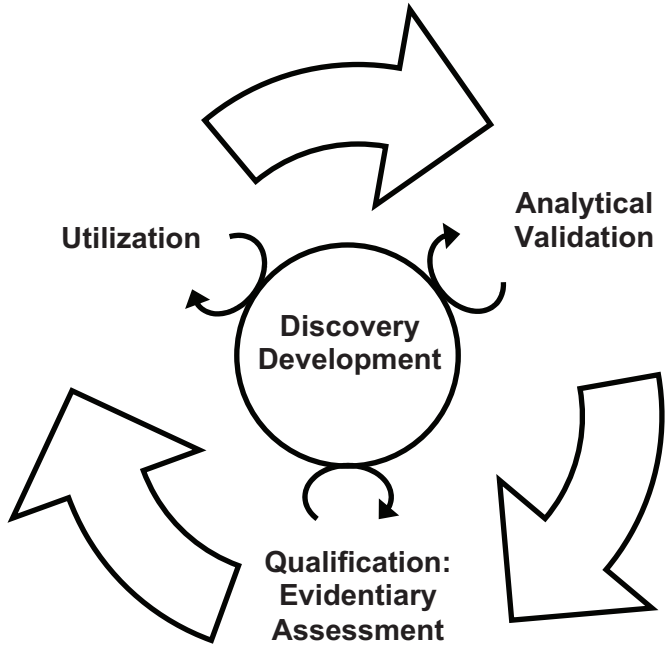

FIGURE S-1 The steps of the evaluation framework are interdependent. While a validated test is required before qualification and utilization can be completed, biomarker uses inform test development, and the evidence suggests possible biomarker uses. In addition, the circle in the center signifies ongoing 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 noninfectious 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 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 of the committee's report (IOM, 2010). 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. 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 (see also Chapters 3 and 4 of the committee's report [IOM, 2010]).

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 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 (see also Chapter 5 of the committee's report [IOM, 2010]).

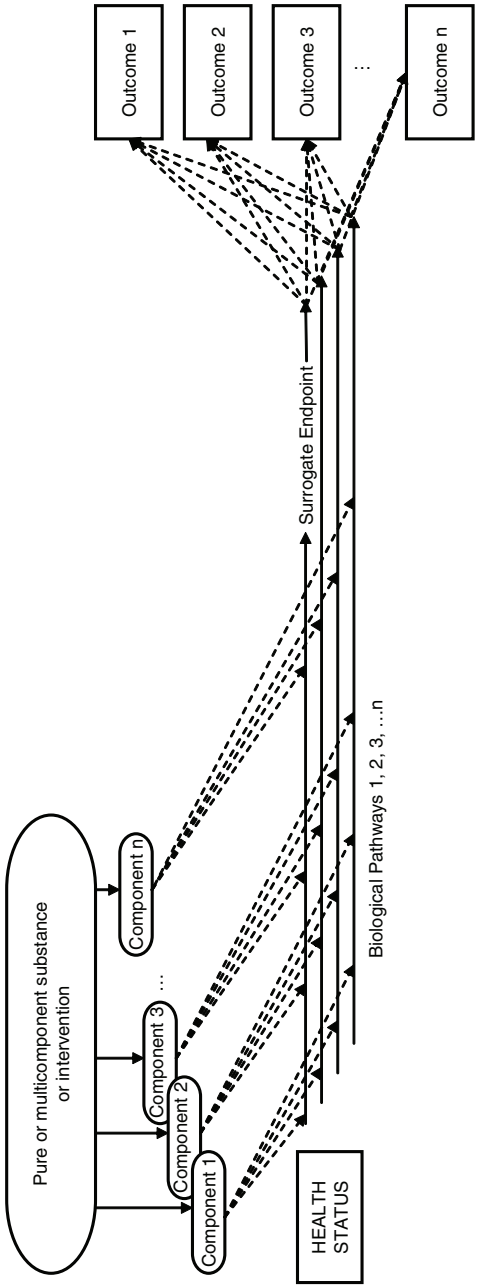

FIGURE S-2 Multiple ingredients, multiple biological pathways, and multiple outcomes illustrate some of the complexities of the use of biomarkers and surrogate endpoints in chronic disease. Note that while the solid horizontal arrows indicate biological pathways, they do not necessarily indicate pathways 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.

IOM. 2010. *Evaluation of biomarkers and surrogate endpoints in chronic disease*. 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.