Free Summary

Evaluation of Biomarkers and Surrogate Endpoints in Chronic Disease

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


This free summary is provided by the National Academies as part of our mission to educate the world on issues of science, engineering, and health. If you are interested in reading the full book, please visit us online at http://www.nap.edu/catalog/12869.html. You may browse and search the full, authoritative version for free; you may also purchase a print or electronic version of the book. If you have questions or just want more information about the books published by the National Academies Press, please contact our customer service department toll-free at 888-624-8373.

This summary plus thousands more available at www.nap.edu.
Summary

Biomarkers are characteristics that are objectively measured and evaluated as indicators of normal biological processes, pathogenic processes, or pharmacologic responses to an intervention. Cholesterol and blood sugar levels are biomarkers, as are blood pressure, enzyme levels, measurements of tumor size from magnetic resonance imaging (MRI) or computed tomography (CT), and the biochemical and genetic variations observed in age-related macular degeneration.

Biomarkers can enable faster, more efficient clinical trials for life-saving and health-promoting interventions. They can help improve understanding of healthy dietary choices, and they can help public health professionals to identify and track health concerns. Biomarkers help health care practitioners and their patients make decisions about patient care. The use of biomarkers depends on the quality of data that supports their use and on the context in which they are applied. Evaluation of the quality of the measurements and data linking the biomarkers to clinical outcomes is important for assessing biomarker utility.

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

BOX S-1 Summary of Recommendations for Effective Biomarker Evaluation

The Evaluation Framework
1. The biomarker evaluation process should consist of the following three steps:
   1a. Analytical validation: analyses of available evidence on the analytical performance of an assay;
   1b. Qualification: assessment of available evidence on associations between the biomarker and disease states, including data showing effects of interventions on both the biomarker and clinical outcomes; and
   1c. Utilization: contextual analysis based on the specific use proposed and the applicability of available evidence to this use. This includes a determination of whether the validation and qualification conducted provide sufficient support for the use proposed.
2a. For biomarkers with regulatory impact, the Food and Drug Administration (FDA) should convene expert panels to evaluate biomarkers and biomarker tests.
2b. Initial evaluation of analytical validation and qualification should be conducted separately from a particular context of use.
2c. The expert panels should reevaluate analytical validation, qualification, and utilization on a continual and a case-by-case basis.
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 simultaneously 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

---

\(^1\) 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).
closely related to patient or consumer experience would be acute myocardial infarction with full recovery and without impact on a patient or consumer’s quality-of-life, and a less clearly related example is an LDL cholesterol level (more accurately, non-HDL cholesterol), as associated with cardiovascular disease mortality. The value of an endpoint increases in relation to the degree to which it conveys information about the effect of an intervention on a patient or consumer’s experience of life. For endpoints that are less clearly related to patient or consumer experience, there is a need to acknowledge that we cannot know with certainty whether a beneficial change in the endpoint will impact a patient or consumer’s experience of life. Further, the committee notes that endpoints can be conceptualized in a spectrum. At one end are endpoints defined by biomarkers alone and 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.

**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.”

Biomarker: “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to an intervention.” 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.” 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.”

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.” Example: blood pressure for trials of several classes of antihypertensive drugs.

**NOTES:**

- The committee defines “objectively” to mean “reliably and accurately.”
- Please see Chapter 2 for discussion of this biomarker.

**SOURCES:**

- Biomarkers Definitions Working Group (2001);
- Wagner (2002); and

**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 diseases.
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 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.

**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.
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 effect of interventions on clinical endpoints of interest; this evidence may also be used to support the associations described in the first component. If the biomarker–clinical endpoint relationship persists over multiple interventions, it is considered more generalizable. It is important to note, however, that the type of reasoning that may be used in qualification is probabilistic rather than deterministic. Although deterministic reasoning ultimately means that every contributing factor to the biomarker–intervention–clinical endpoint link is defined and understood, probabilistic reasoning emphasizes epidemiological and statistical relationships, acknowledging that all contributing factors are generally not fully understood and that some factors may be fundamentally random.

Related to the first component of qualification, prognostic value can be assessed by using concepts described by criteria proposed for establishing causation of non-infectious diseases (Advisory Committee to the Surgeon General, 1964; Hill, 1965). These criteria evaluate characteristics such as temporality, strength of association, biological plausibility, and consistency, among others. Given that biomarkers are “indicators”—in that they are not necessarily causal—and that an abnormal value or a gradient in level over time is not necessarily informative or predictive depending on the clinical situation, the committee instead used these criteria as a structure for assessing the prognostic value, or degree of association between the biomarker and the clinical outcomes of interest absent any interventions. For a surrogate endpoint, or a biomarker deemed useful as a substitute for a defined, disease-relevant clinical endpoint, prognostic value is a necessary—but not sufficient—criterion for the evaluation. Depending on the situation, not all of the criteria must be fulfilled; temporality, strength of association, and consistency are particularly important, however. Observational data in human populations and preliminary clinical data (e.g., phase I or II data) are considered. Nonetheless, determination of whether a biomarker can be used as a surrogate endpoint for a specified intervention is done in the utilization step of the evaluation process.

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

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

Strong evidence and a compelling context are needed for the utilization of a biomarker as a surrogate endpoint in situations with regulatory impact. In the case of chronic disease, where there are multiple pathogenetic pathways leading to development of clinical outcomes and multiple manifestations of disease, the probabilistic nature of predictions made using biomarker data means that no biomarker can give absolute certainty of an event’s future occurrence nor absolute certainty of the timing of the predicted event. Nonetheless, there are situations in which use of a biomarker as a surrogate endpoint in situations with regulatory impact may be supported, such as in situations where the need for interventions is urgent or where studies including clinical endpoints are not feasible because of technical or ethical reasons. Situations with regulatory impact are defined in chapter 3. Again, this is not meant to discourage use of biomarkers in product development; biomarkers play an important role in research and decision-making. Finally, it is essential to remember that the information that an individual surrogate endpoint or clinical endpoint can give is inherently limited; as a result, it is important to emphasize the need to evaluate data relating to adverse events and unintended effects of biomarker use. As will be discussed and shown in chapters 3 and 4, the status of a biomarker as a surrogate endpoint is context-specific, and a biomarker cannot be assumed to be a general surrogate endpoint separate from a designated use.

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

**Scientific Process Harmonization**

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

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

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

**Recommendation 4:**
The FDA should take into account a nutrient’s or food’s source as well as any modifying effects of the food or supplement that serves as the delivery vehicle and the dietary patterns associated with consumption of the nutrient or food when reviewing health-related label claims and the safety of food and supplements.

Drugs, biologics, and devices are evaluated for efficacy and safety on the basis of the whole products. Recommendation 4 seeks to extend this approach to foods and supplements. Due to the importance of the differing health effects of individual nutrients or other food substances in food or supplement products composed of multiple substances, 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.

**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.

**Ancillary Recommendations**

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

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

Improving Evidence-Based Regulation

5a. Congress should strengthen the FDA’s authority to request and enforce post-market 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


Evaluation of Biomarkers and Surrogate Endpoints in Chronic Disease

Committee on Qualification of Biomarkers and Surrogate Endpoints in Chronic Disease
Board on Health Care Services
Board on Health Sciences Policy
Food and Nutrition Board

Christine M. Micheel and John R. Ball, Editors

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

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

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

International Standard Book Number 0-309-XXXXX-X (Book)
International Standard Book Number 0-309-XXXXX-X (PDF)
Library of Congress Control Number: 00 XXXXXX

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.

Copyright 2010 by the National Academy of Sciences. All rights reserved.

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.

“Knowing is not enough; we must apply. Willing is not enough; we must do.”

—Goethe
COMMITTEE ON QUALIFICATION OF BIOMARKERS AND SURROGATE ENDPOINTS IN CHRONIC DISEASE

JOHN R. BALL (Chair), Executive Vice President, American Society for Clinical Pathology
MICHELLE A. ALBERT, Assistant Professor of Medicine, Associate Physician, and Director of Behavioral and Neurocardiovascular Cardiology, Brigham and Women’s Hospital, Harvard Medical School
FRED APPLE, Medical Director, Clinical Laboratories, Hennepin County Medical Center, and Professor, Laboratory Medicine and Pathology, University of Minnesota School of Medicine
ROBERT M. CALIFF, Vice Chancellor for Clinical Research and Professor of Medicine, Cardiology, Duke University School of Medicine
VICTOR DE GRUTTOLA, Chair and Professor, Biostatistics, Harvard School of Public Health
DAVID DEMETS, Professor and Chair, Biostatistics & Medical Informatics, University of Wisconsin–Madison
ROBERT GERSZTEN, Research Director and Faculty Member, Massachusetts General Hospital, and Associate Professor of Medicine, Harvard Medical School
WILLIAM 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, Associate Professor, Physiology and Biochemistry, Johns Hopkins University
JOHN A. WAGNER, Vice President, Clinical Pharmacology, Merck and Company, 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, Research Associate
BERNADETTE MCFADDEN, Research Associate (from July 2009 to January 2010)
LISA BOYETTE, Christine Mirzayan Science and Technology Policy Graduate Fellow (September to November 2009)
ANNA WOLOSZYNSKA-READ, Christine Mirzayan Science and Technology Policy Graduate Fellow (January to April 2009)
CAIRA WOODS, Christine Mirzayan Science and Technology Policy Graduate Fellow (January to April 2010)
DESH MOHAN, Intern (June to August 2009)
ASHLEY MCWILLIAMS, Senior Program Assistant
PATRICK BURKE, Financial Associate
ROGER HERDMAN, Director, Board on Health Care Services
LINDA MEYERS, Director, Food and Nutrition Board
ANDREW POPE, Director, Board on Health Sciences Policy
REVIEWERS

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

**SUMMARY**
- Study Scope, S-3
- Findings, Conclusions, and Recommendations, S-4
- References, S-12

**1 INTRODUCTION**
- Origin of the Task, 1-2
- Definitions, 1-4
- Related IOM Work, 1-10
- Framework of the Report, 1-11
- References, 1-12

**2 REVIEW: EVALUATING AND REGULATING BIOMARKER USE**
- Introduction, 2-1
- Survey of Biomarker Uses, 2-1
- Evaluation Frameworks, 2-15
- Evolution of Regulatory Perspectives on Surrogate Endpoints, 2-24
- Biomarkers and Communication Strategies at the FDA, 2-35
- Further Issues with Use of Biomarkers, 2-36
- References, 2-40

**3 THE BIOMARKER EVALUATION PROCESS**
- The Rationale for an Interrelated, Three-Step Process, 3-3
- Application of the Evaluation Framework, 3-16
- Scientific Process Harmonization, 3-19
- Conclusion, 3-22
- References, 3-23

**4 CASE STUDIES**
- Introduction, 4-1
- Tumor Size as Biomarker for Cancer Clinical Endpoints, 4-3
- C-Reactive Protein, 4-8
- Troponin, 4-16
- LDL and HDL as Biomarkers for Cardiovascular Risk, 4-21
- Beta-Carotene, 4-28
- References, 4-34

**5 STRENGTHENING EVIDENCE-BASED REGULATION**
- Chapter Recommendations, 5-1
- FDA Regulatory Authority, 5-2
- Federal Agencies and Data Collection, 5-17
- Tracking the Effects of Biomarker Use at the FDA, 5-22
- References, 5-27
BOXES, FIGURES, AND TABLES

Summary
Boxes
S-1 Summary of Recommendations for Effective Biomarker Evaluation, S-1
S-2 Important Definitions, S-3
S-3 Ancillary Recommendations, S-11

Figures
S-1 Steps of the Evaluation Framework, S-5
S-2 Complexities of Biomarker Use, S-10

Chapter 1
Box
1-1 Important Definitions, 1-5

Tables
1-1 Use of Biomarkers in Chronic Disease Patient Care, 1-6
1-2 Use of Biomarkers in Drug Development, 1-6
1-3 Regulatory Definitions of Surrogate Endpoint, 1-7
1-4 Literature Definitions of Surrogate Endpoint, 1-8

Chapter 2
Boxes
2-1 Characteristics of Comparative Effectiveness Research, 2-5
2-2 Hill’s Criteria, 2-16
2-3 Five Unique Features of Decision Analysis, 2-20
2-4 FDA’s Risk Communication Advisory Committee, 2-26

Figures
2-1 Reasons for Failure of Surrogate Endpoints, 2-12
2-2 Setting for Greatest Potential for Useful Surrogate Endpoint, 2-12
2-3 Receiver Operative Characteristic Graph, 2-21
2-4 Biomarker Qualification Pilot Process (FDA), 2-28
2-5 FDA Ranking System for Health Claims, 2-33

Tables
2-1 Categories of Biomarker Use, 2-2
2-2 Proposed Evidence Map for Biomarker Qualification (Altar et al., 2008), 2-23
2-3 List of Regulations and Guidances Pertaining to Surrogate Endpoints, 2-25
2-4 Health Claims Based on Surrogate Endpoints, 2-31
2-5 Qualified Health Claims Approved by the FDA, 2-33

Chapter 3
Boxes
3-1 Recommendations 1–4, 3-1
3-2 Tumor Size and Analytical Validation, 3-8
3-3 CRP, Inflammatory Markers, and Qualification, 3-11
3-4 Troponin and Utilization, 3-14
3-5 LDL and HDL Cholesterol and Surrogacy, 3-16
3-6 Blood Levels of Beta-Carotene, 3-19

Figure
3-1 Steps of the Evaluation Framework, 3-2
Tables
3-1 Sources of Variability in Biomarker Measurements, 3-6
3-2 Information Needed for Package Inserts and Peer-Reviewed Publications Describing Biomarker Assays, 3-7
3-3 Utilization: Critical and Important Factors for Consideration, 3-12

Chapter 4
Box
4-1 Conditions Associated with High Cardiac Troponins, 4-19

Figure
4-1 Inflammatory Risk Factors, 4-10

Tables
4-1 Brief Summary of the Results of the Case Studies, 4-1
4-2 Assays of Inflammatory Markers for Potential Clinical Use, 4-11

Chapter 5
Boxes
5-1 Expanding FDA Responsibilities, 5-3
5-2 Core FDA Regulatory Functions, 5-5
5-3 Dietary Supplements, 5-12
5-4 Role of NIH in Biomarker Data Collection, 5-22

Figures
5-1 FDA—Regulatory Industry (2006), 5-5
5-2 FDA Health Claims and Dietary Guidance Statements, 5-15

Table
5-1 Types of Health Claims, 5-10

Appendix A
Tables
A-1 Historical Review of Biomarker–Surrogate Endpoint Literature, A-1
A-2 Continuation of Table A-1 for 2007–2009, A-11

Appendix B
Boxes
B-1 Summary of Recommendations to Develop Biomarker-Based Tools for Cancer, B-1
B-2 Summary of Recommendations for The Future of Drug Safety: Promoting and Protecting the Health of the Public, B-2