



Generating Evidence for Genomic Diagnostic Test Development: Workshop Summary

ISBN
978-0-309-21104-8

100 pages
6 x 9
PAPERBACK (2011)

Theresa Wizemann and Adam C. Berger, Rapporteurs; Roundtable on Translating Genomic-Based Research for Health; Institute of Medicine

 Add book to cart

 Find similar titles

 Share this PDF



Visit the National Academies Press online and register for...

- ✓ Instant access to free PDF downloads of titles from the
 - NATIONAL ACADEMY OF SCIENCES
 - NATIONAL ACADEMY OF ENGINEERING
 - INSTITUTE OF MEDICINE
 - NATIONAL RESEARCH COUNCIL
- ✓ 10% off print titles
- ✓ Custom notification of new releases in your field of interest
- ✓ Special offers and discounts

Distribution, posting, or copying of this PDF is strictly prohibited without written permission of the National Academies Press. Unless otherwise indicated, all materials in this PDF are copyrighted by the National Academy of Sciences. Request reprint permission for this book

GENERATING EVIDENCE FOR GENOMIC DIAGNOSTIC TEST DEVELOPMENT

WORKSHOP SUMMARY

Theresa Wizemann and Adam C. Berger, *Rapporteurs*

Roundtable on Translating Genomic-Based Research for Health

Board on Health Sciences Policy

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 project was supported by contracts between the National Academy of Sciences and the American College of Medical Genetics (unnumbered contract); American Medical Association (unnumbered contract); American Nurses Association (unnumbered contract); Blue Cross and Blue Shield Association (unnumbered contract); Centers for Disease Control and Prevention (Contract No. 200-2005-13434); College of American Pathologists (unnumbered contract); Department of the Air Force (Contract No. FA7014-10-P-0072); Department of Veterans Affairs (Contract No. V101(93) P-2238); Eli Lilly and Company (Contract No. LRL-0028-07); Genetic Alliance (unnumbered contract); Health Resources and Services Administration; Johnson & Johnson (unnumbered contract); Kaiser Permanente (unnumbered contract); National Cancer Institute (Contract No. N01-OD-4-2139, TO#189); National Heart, Lung, and Blood Institute (Contract No. N01-OD-4-2139, TO#189); National Human Genome Research Institute (Contract No. N01-OD-4-2139, TO#189); National Society of Genetic Counselors (unnumbered contract); and Pfizer Inc. (Contract No. 140-N-1818071). Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the organizations or agencies that provided support for the project.

International Standard Book Number-13: 978-0-309-21104-8

International Standard Book Number-10: 0-309-21104-2

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.

Copyright 2011 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.

Suggested citation: IOM (Institute of Medicine). 2011. *Generating Evidence for Genomic Diagnostic Test Development: Workshop 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.

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

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

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

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

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

www.national-academies.org

PLANNING COMMITTEE¹

- NAOMI ARONSON**, Executive Director, Technology Evaluation Center, Blue Cross and Blue Shield Association, Chicago, IL
- ANDREW FREEDMAN**, Branch Chief, Clinical and Translational Epidemiology Branch, Epidemiology and Genetics Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Rockville, MD
- ROGER D. KLEIN**, Director, Molecular Oncology Laboratory at BloodCenter of Wisconsin; Clinical Assistant Professor of Pathology, Medical College of Wisconsin, Milwaukee
- DEBRA LEONARD**, Professor and Vice Chair for Laboratory Medicine, Director of the Clinical Laboratories, and Director of the Pathology Residency Training Program, Weill Cornell Medical Center of Cornell University, New York, NY
- ELIZABETH MANSFIELD**, Director of the Personalized Medicine Staff, Office of In Vitro Diagnostic Device Evaluation and Safety in the Center for Devices and Radiological Health, Food and Drug Administration, Silver Spring, MD
- ROBERT McCORMACK**, Director for Technology Assessment, Ortho-Clinical Diagnostics; Vice President of Scientific and Medical Affairs, Veridex, LLC, Raritan, NJ
- ROBERT NUSSBAUM**, Chief of the Division of Medical Genetics, Department of Medicine, Institute of Human Genetics, San Francisco, CA
- KIM POPOVITS**, President and Chief Executive Officer, Genomic Health, Redwood City, CA
- RONALD PRZYGODZKI**, Associate Director for Genomic Medicine, Acting Director of Biomedical Laboratory Research and Development, Office of Research and Development, Department of Veterans Affairs, Washington, DC
- ALLEN ROSES**, President and Chief Operating Officer, Cabernet, Shiraz and Zinfandel Pharmaceuticals; Jefferson–Pilot Professor of Neurobiology and Genetics, Professor of Medicine (Neurology), Director of the Deane Drug Discovery Institute; Senior Scholar at the Fuqua School of Business, Duke University, Durham, NC
- DANIEL WATTENDORF**, Deputy Chief, Medical Innovations, Department of the Air Force; Program Manager, DARPA/Defense Sciences Office, Arlington, VA

¹ Institute of Medicine planning committees are solely responsible for organizing the workshop, identifying topics, and choosing speakers. The responsibility for the published workshop summary rests with the workshop rapporteurs and the institution.

IOM Staff

ADAM C. BERGER, Project Director

CLAIRE F. GIAMMARIA, Research Associate

TONIA E. DICKERSON, Senior Program Assistant

ROUNDTABLE ON TRANSLATING GENOMIC-BASED RESEARCH FOR HEALTH¹

- WYLIE BURKE** (*Chair*), Professor and Chair, Department of Bioethics and Humanities, University of Washington, Seattle
- NAOMI ARONSON**, Executive Director, Technology Evaluation Center, Blue Cross and Blue Shield Association, Chicago, IL
- BRUCE BLUMBERG**, Co-Chief of Medical Genetics, Kaiser Permanente, and Institutional Director of Graduate Medical Education, Northern California Kaiser Permanente, The Permanente Medical Group, Oakland, CA
- DENISE E. BONDS**, Medical Officer, Division of Prevention and Population Sciences, National Heart, Lung, and Blood Institute, Bethesda, MD
- C. THOMAS CASKEY**, Director and Chief Executive Officer of the Brown Foundation Institute of Molecular Medicine, the George & Cynthia Mitchell Distinguished Chair in Neurosciences, and Executive Vice President of Molecular Medicine and Genetics, University of Texas Health Science Center at Houston
- STEPHEN ECK**, Vice President, Translational Medicine and Pharmacogenomics, Eli Lilly and Company, Indianapolis, IN
- W. GREGORY FEERO**, Special Advisor to the Director for Genomic Medicine, National Human Genome Research Institute, Bethesda, MD
- ANDREW N. FREEDMAN**, Branch Chief, Clinical and Translational Epidemiology Branch, Epidemiology and Genetics Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Rockville, MD
- GEOFFREY GINSBURG**, Director, Center for Genomic Medicine, Institute for Genomic Sciences and Policy, Duke University, Durham, NC
- SHARON KARDIA**, Professor and Chair of Epidemiology; Director, Public Health Genetics Program; Director, Life Science and Society Program; Co-Director, Center for Genomics & Public Health, University of Michigan School of Public Health, Ann Arbor
- MOHAMED KHAN**, representing the American Medical Association; Leader of Radiation Oncology, Vancouver Cancer Centre, BC Cancer Agency, Vancouver, BC, Canada

¹ Institute of Medicine Forums and Roundtables do not issue, review, or approve individual documents. The responsibility for the published workshop summary rests with the workshop rapporteurs and the institution.

- MUIN KHOURY**, Director, National Office of Public Health Genomics, Centers for Disease Control and Prevention, Atlanta, GA
- DEBRA LEONARD**, representing the College of American Pathologists; Professor and Vice Chair for Laboratory Medicine, Director of the Clinical Laboratories, and Director of the Pathology Residency Training Program, Weill Cornell Medical Center of Cornell University, New York, NY
- MICHELE LLOYD-PURYEAR**, Chief, Genetic Services Branch, Health Resources and Services Administration, Rockville, MD
- ELIZABETH MANSFIELD**, Director of the Personalized Medicine Staff, Office of In Vitro Diagnostic Device Evaluation and Safety in the Center for Devices and Radiological Health, Food and Drug Administration, Silver Spring, MD
- GARRY NEIL**, Corporate Vice President, Corporate Office of Science and Technology, Johnson & Johnson, New Brunswick, NJ
- ROBERT L. NUSSBAUM**, Chief, Division of Medical Genetics, Department of Medicine and Institute of Human Genetics, University of California–San Francisco School of Medicine
- AIDAN POWER**, Vice President and Global Head of Molecular Medicine, Pfizer Inc., Groton, CT
- RONALD PRZYGODZKI**, Associate Director for Genomic Medicine, Biomedical Laboratory Research and Development, Department of Veterans Affairs, Washington, DC
- ALLEN D. ROSES**, President and Chief Operating Officer, Cabernet, Shiraz and Zinfandel Pharmaceuticals; Jefferson–Pilot Professor of Neurobiology and Genetics, Professor of Medicine (Neurology); Director, Deane Drug Discovery Institute; Senior Scholar, Fuqua School of Business, R. David Thomas Executive Training Center, Duke University, Durham, NC
- KEVIN A. SCHULMAN**, Professor of Medicine and Business Administration; Director, Center for Clinical and Genetic Economics; Associate Director, Duke Clinical Research Institute, Duke University School of Medicine, Durham, NC
- SHARON TERRY**, President and Chief Executive Officer, Genetic Alliance, Washington, DC
- MARTHA TURNER**, Assistant Director, American Nurses Association Center for Ethics and Human Rights, Silver Spring, MD
- MICHAEL S. WATSON**, Executive Director, American College of Medical Genetics, Bethesda, MD
- DANIEL WATTENDORF**, Deputy Chief, Medical Innovations, Department of the Air Force; Program Manager, DARPA/Defense Sciences Office, Arlington, VA

CATHERINE A. WICKLUND, Past President, National Society of Genetic Counselors; Director, Graduate Program in Genetic Counseling; Assistant Professor, Department of Obstetrics and Gynecology, Northwestern University, Chicago, IL

IOM Staff

ADAM C. BERGER, Project Director

CLAIRE F. GIAMMARIA, Research Associate

TONIA E. DICKERSON, Senior Program Assistant

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:

Margaret Piper, Director of Genomics Resources, Blue Cross and Blue Shield Association Technology Evaluation Center, Atlanta, GA

Jeffrey Roche, Medical Officer, Office of Clinical Standards and Quality, Center for Medicare and Medicaid Services, Baltimore, MD

Steven Shak, Chief Medical Officer, Genomic Health, Inc., Redwood City, CA

Sean Tunis, Director, Center for Medical Technology Policy, Baltimore, MD

Although the reviewers listed above have provided many constructive comments and suggestions, they did not endorse the final draft of the report before its release. The review of this report was overseen by **Harold J.**

Fallon, Dean Emeritus of the University of Alabama at Birmingham School of Medicine. Appointed by the Institute of Medicine, he 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 author and the institution.

Acknowledgments

The support of the sponsors of the Institute of Medicine Roundtable on Translating Genomic-Based Research for Health was crucial to the planning and conduct of the workshop “Generating Evidence for Genomic Diagnostic Test Development.” Federal sponsors are the Department of the Air Force; Centers for Disease Control and Prevention; Health Resources and Services Administration; National Cancer Institute; National Heart, Lung, and Blood Institute; National Human Genome Research Institute; and the Department of Veterans Affairs. Non-federal sponsorship was provided by the American College of Medical Genetics, the American Medical Association, the American Nurses Association, Blue Cross and Blue Shield Association, the College of American Pathologists, Eli Lilly and Company, Genetic Alliance, Johnson & Johnson, Kaiser Permanente, the National Society of Genetic Counselors, and Pfizer Inc.

The Roundtable wishes to express its gratitude to the expert speakers whose presentations helped outline the challenges and opportunities in generating evidence for genomic diagnostic tests. The Roundtable also wishes to thank the members of the planning committee for their work in developing an excellent workshop agenda.

Contents

ABBREVIATIONS AND ACRONYMS	xix
1 INTRODUCTION	1
2 STAKEHOLDER PERSPECTIVES ON EVIDENCE	5
FDA Perspective, 7	
Payer Perspective, 10	
Evidence-Based Review Group Perspective, 14	
Health-Care Provider Perspective, 19	
Panel Discussion, 22	
3 APPROACHES TO EVIDENCE GENERATION	25
Pharmacogenomics Clinical Trials, 25	
Use of Archived Samples to Evaluate Genomic Tests, 31	
Coverage with Evidence Development, 33	
Constructing Chains of Evidence, 39	
Discussion, 41	
4 OVERCOMING BARRIERS FOR EVIDENCE GENERATION	43
Balancing Stakeholder Needs, 43	
Public-Private Partnerships, 44	
The Evidentiary Bar for Clinical Utility, 46	
Data Sources, 46	

5	CONSIDERATIONS MOVING FORWARD	53
	Medical Practice Versus Evidence-Based Recommendations, 53	
	Heterogeneity of Evidence-Based Decisions, 54	
	Research, 56	
	Broad Considerations, 57	
	Transparency, 57	
	Public Expectations, 58	
6	FINAL REMARKS	59
	Chair’s Summary, 60	
	REFERENCES	63
	APPENDIXES	
A	WORKSHOP AGENDA	67
B	SPEAKER BIOGRAPHICAL SKETCHES	71
C	REGISTERED ATTENDEES	81

Tables, Figures, and Boxes

TABLES

- 1-1 Evidence-Based Review of Select Genomic Tests, 3
- 2-1 Oncotype Dx Reclassification of Patients, 13
- 3-1 Classic Randomized Controlled Trials (RCTs) Versus Practical Clinical Trials (PCTs), 28
- 3-2 Summary of Ontario Field Evaluations, 36

FIGURES

- 2-1 Sample analytic framework, 16
- 2-2 Breast cancer mortality rates from 1950-2003 and the effect of screening and adjuvant therapy, 20
- 3-1 Pharmacogenomic trial designs, including retrospective, prospective stratified, and prospective screened, 26
- 3-2 Hypothetical genotype-guided versus usual care scenario to measure smoking cessation rates using varenicline or patch, 30
- 3-3 Prospective stratified RCT scheme, 30
- 3-4 MAS- and OHTAC-associated structures and linkages, 34

BOXES

- 1-1 Speakers' Perspectives on Barriers to the Collection of Clinical Validity and Utility Data for Genomic Diagnostic Tests, 4
- 2-1 Stakeholder Perspectives on Evidence, 6
- 2-2 Challenges to Establishing Analytical and Clinical Validity of Diagnostics, 8

Abbreviations and Acronyms

ACCE	analytic validity, clinical validity, clinical utility, and associated ethical, legal, and social implications model
AHRQ	Agency for Healthcare Research and Quality
ASCO	American Society of Clinical Oncology
CAP	College of American Pathologists
CDC	Centers for Disease Control and Prevention
CDRH	Center for Devices and Radiological Health (FDA)
CLIA	Clinical Laboratory Improvement Amendments
EGAPP	Evaluation of Genomic Applications in Practice and Prevention
EGFR	epidermal growth factor receptor
eMERGE	Electronic Medical Records and Genomics
ER	estrogen receptor
FDA	U.S. Food and Drug Administration
GAPPNet	Genomic Applications in Practice and Prevention Network
GWAS	genome-wide association study
HMO	health maintenance organization
IHC	immunohistochemistry

LOE	level of evidence
MAQC	microarray quality control
MAS	Medical Advisory Secretariat
MSI	microsatellite instability
NCI	National Cancer Institute (NIH)
NHGRI	National Human Genome Research Institute
NIDDK	National Institute of Diabetes and Digestive and Kidney Diseases (NIH)
NIH	National Institutes of Health
OHTAC	Ontario Health Technology Advisory Committee
PET	positron emission tomography
PSA	prostate-specific antigen
RCT	randomized controlled trial
SSRI	selective serotonin reuptake inhibitor
TEC	Technology Evaluation Center (Blue Cross and Blue Shield Association)

1

Introduction

The field of genomics has expanded greatly since the first sequence of the human genome was published a decade ago. According to workshop chair Debra Leonard of Weill Cornell Medical College, the hoped-for outcomes of the human genome project were to understand human genetic variations and their relationship to health and disease; to predict disease risks for prevention and earlier treatment; to refine disease diagnosis by understanding the underlying genetic variance and molecular mechanisms; and to use that information to create better treatments and improve the health and health outcomes of the U.S. population. Over the past 10 years, scientists have developed genomic tests based on identified gene-disease associations that can predict the response of an individual patient to a drug intervention or the risk of developing Alzheimer's disease. However, much of the evidence surrounding the clinical value and utility of these tests has not been sufficient enough for clinical practitioners to broadly embrace many of these in practice.

A major impediment to the integration of these genomic tests into routine health care is the lack of an adequate evidence base linking the use of genomic diagnostic tests to health outcomes. Since these new technologies are beginning to play an increased role in clinical decision-making and the management of disease, the Institute of Medicine's Roundtable on Translating Genomic-Based Research for Health hosted a public workshop on

November 17, 2010, to explore issues related to this lack of evidence.¹ Various stakeholders, including regulators and policymakers, payers, health-care providers, researchers, funders, and evidence-based review groups, were invited to share their perspectives on the strengths and limitations of the evidence being generated to assess the clinical validity and utility of genomic diagnostic tests. Specifically, panelists were asked to address the following:

- What evidence is required by stakeholders (e.g., for decisions regarding clearance, use, and reimbursement)?
- How is evidence currently being generated?
- Are there innovative and efficient ways to generate high-quality evidence?
- How can the barriers to generating this evidence be overcome?

Early genetic tests, Leonard explained, were focused on single genes. The market was limited and reimbursement was poor, and the *in vitro* diagnostics industry was therefore not very interested in developing genetic tests. Instead, genetic tests for the diagnosis of disease were generally developed as needed by clinical laboratories. These were based on published genotype–phenotype correlations and were developed using standard molecular biology methods and sets of patient and control samples. The Clinical Laboratory Improvement Amendments (CLIA) (42 U.S.C. 263a) allows such practices without the need for receiving device clearance from the U.S. Food and Drug Administration (FDA). However, Leonard said, there were concerns about the quality of these tests, the potential for harm to patients, the clinical validity and utility of the tests, and the relatively expensive cost.

Genetic tests are still in use today, Leonard said, but the focus has shifted to *genomic tests*, which are complex testing algorithms of multiple genetic variants, multiple genes, or gene expression patterns. Genomic tests are used for diagnosis as well as for therapeutic selection, dosing, prognosis, and residual disease detection. However, the majority of these tests have insufficient clinical validity and utility data, and there is currently little evidence of improved health outcomes from their use (Table 1-1). The increasing role of genomic tests in clinical decision-making has led to

¹ This workshop was organized by an independent planning committee whose role was limited to developing the meeting agenda. This summary has been prepared by the rapporteurs as a factual summary of the discussion that took place at the workshop. All views presented in the report are those of the individual workshop participants and should not be construed as reflecting any group consensus.

INTRODUCTION

TABLE 1-1 Evidence-Based Review of Select Genomic Tests

Genetic Test	Reviewed by	Conclusion
Thrombophilia tests	AHRQ/ EGAPP	No direct evidence for improved outcomes
HER2 testing in breast cancer	AHRQ	Weak evidence relating test result to treatment outcomes
Gene expression profiles for breast cancer	AHRQ	High quality retrospective clinical utility data for Oncotype DX
UGT1A1 genotyping for colorectal cancer patients	EGAPP	Insufficient evidence for or against testing
Genetic testing for hereditary nonpolyposis colorectal cancer (HNPCC)	EGAPP	Limited evidence that mismatch repair gene mutations cause family members to have increased screening
CYP450 for non-psychotic depression	EGAPP	Paucity of good quality evidence that testing is useful for selective serotonin reuptake inhibitor (SSRI) outcomes
Genomic tests for ovarian cancer	EGAPP	No evidence that tests affect outcomes in asymptomatic women

Abbreviations: AHRQ (Agency for Healthcare Research and Quality), EGAPP (Evaluation of Genomic Applications in Practice and Prevention).

SOURCE: Adapted from Leonard, IOM workshop presentation on November 17, 2010.

ongoing discussions at FDA regarding the appropriate level of regulatory oversight for genetic and genomic tests.²

The barriers to evidence generation have been discussed in many venues (summarized in Box 1-1). The goal of this workshop, Leonard said, is to look beyond these barriers to define the evidence needed and the mecha-

² As further discussed in Chapter 2, other stakeholder groups including payers, evidence-based review groups, providers, professional societies, and patient groups have also initiated discussions on the utility of genetic and genomic tests in clinical decision making.

BOX 1-1**Speaker's Perspectives on Barriers to the Collection of Clinical Validity and Utility Data for Genomic Diagnostic Tests**

- Various stakeholders require different types and levels of evidence (e.g., doctors, patients, FDA, payers, evidence-based review groups)
- Limited or nonexistent funding for randomized controlled trials of genomic tests
- Length of time needed for clinical trials to be completed
- High cost of archiving specimens from therapeutic clinical trials
- Lack of access to annotated clinical specimens

nisms to obtain it so that the promise of the human genome project and genomic diagnostic testing can be fully realized.

The report that follows summarizes the presentations and discussions by the expert panelists. Chapter 2 provides the different stakeholder perspectives on the type and level of evidence needed for decision making. Approaches for evidence generation are discussed in Chapter 3. Chapters 4 and 5 examine ways to overcome the barriers to evidence generation and strategies for moving forward. Final remarks are provided in Chapter 6, and the workshop agenda, biographical sketches of the panelists, and list of attendees are included in the appendixes.

2

Stakeholder Perspectives on Evidence

Key Points Raised by Speakers

- More dialogue and coordination among stakeholders is needed to facilitate the development of the necessary evidence base.
- Test development and reimbursement need to focus on the clinical utility of the test and the net benefit to patients.
- The analysis of evidence must be adapted to the clinical setting and to the evidence needed for that particular application.

Personalized medicine has multiple stakeholders, including regulators and policymakers; evidence-based review groups; health-care providers; payers; academic, industry, and government researchers and developers; and patients. Enabling the validation and utilization of genomic-based diagnostic tests involves understanding the views of these interested parties. Stakeholders representing regulators, payers, evidence-based review groups, and providers shared their perspectives on evidence (summarized in Box 2-1), illustrated through case examples.

BOX 2-1

Stakeholder Perspectives on Evidence

FDA

- Regulatory decision to approve or clear a diagnostic device for marketing
- Focused on safety and effectiveness
 - Analytical validity: sufficiently accurate and precise measurement of the analyte
 - Clinical validity: the biological and medical significance of the test result
- Risk-based classification approach to regulation of devices
- More complex, higher-risk devices require more substantial levels of evidence
- Bound by law to declare or decline approval/clearance based on evidence submitted by product sponsor (i.e., no mechanism for provisional approval)

Payers

- Insurer decision to cover and reimburse the use of a diagnostic device
- Focused on outcomes
- Clinical utility: the impact of the test on patient management and outcomes compared with usual care
- May choose to provisionally cover a test pending collection of further evidence of utility (coverage with evidence development)

Evidence-Based Review Groups

- Independent review of evidence which, in the case of EGAPP or medical professional organizations, can result in practice recommendations
- Focused on net benefit
 - Balance of benefits or potential benefits with harms or potential harms
- Recommendations
 - For or against the use of a product based on quality of the evidence and certainty of net benefit
 - No recommendation made in cases of insufficient evidence
 - Applications are further classified as neutral, discouraging, or encouraging based on contextual factors
 - Products with insufficient but encouraging evidence may merit limited use as additional data are gathered

Providers

- Focused on value to the individual patient
- Increase chances of cure/survival/palliation
- Decrease exposure to toxicity from unnecessary or inappropriate therapy
- Decision to treat differs from patient to patient
- Affected by patient, provider, and societal perspectives on risks, benefits, and costs
- Two patients given the same test result may come to different treatment decisions (regardless of practice recommendations)

FDA PERSPECTIVE

A Focus on Safety and Effectiveness

The core function of the Center for Devices and Radiological Health (CDRH) at the U.S. Food and Drug Administration (FDA) is the review and clearance or approval of medical devices for marketing, said Robert Becker of the Office of In Vitro Diagnostic Device Evaluation and Safety, and such clearance or approval is based on evaluations of safety and effectiveness. Regulations mandate that for safety there must be a reasonable assurance that the probable benefits outweigh any probable risks [21CFR860.7(d)(1)] and that for effectiveness there is a reasonable assurance that the use of the device will provide clinically significant results [21CFR860.7(e)(1)].

FDA classifies in vitro diagnostic devices on the basis of risk. Class III devices are the most complex and present the highest risk (e.g., cancer diagnostics). Risk in this case, Becker clarified, refers to the potential consequences of an inaccurate test result, not to risks associated with the testing process itself. Makers of Class III devices must submit a premarket approval application demonstrating safety and effectiveness and obtain FDA approval prior to marketing. Moderate-risk Class II devices, such as those used to monitor a patient already diagnosed with cancer, are subject to premarket notification [also known as the “510(k) process”], in which the sponsor must demonstrate substantial equivalence of the device to an already marketed product. Class I devices are common, low-risk devices which are generally exempt from premarket evaluation by the agency. The regulatory aims for in vitro diagnostics, Becker said, are clarity and reliability concerning the test description, its intended use, the instructions for use, performance claims, manufacturing, and detection and resolution of issues that arise after the test is on the market.

Performance Claims

The performance claims and risk-based classification of a diagnostic test are based on how well the test supports the intended use. FDA requires evidence of *analytical validity*, the accuracy in and reliability of measuring the analyte of interest, and *clinical validity*, the accuracy and reliability of the test in identifying or predicting the biological and medical significance of the test result. The focus is on safety and effectiveness, and in its general review of devices the agency does not address clinical utility (the impact on patient care and outcomes), costs, or comparative effectiveness. The FDA does keep in mind how the tests are going to be used and whether they will be able to effectively guide medical care, Becker said, although this is not part of the official review process.

BOX 2-2
**Challenges to Establishing Analytical and
Clinical Validity of Diagnostics**

Analytical Validity

- Availability of reference methods and materials
- Analytical specifications for multivariate tests
- Clinical samples
- Number, kind, age, and storage
- Full-spectrum assessment
- Pre-analytical, complete analytical process

Clinical Validity

- Sufficient number of patients and samples
- Rare alleles, private variants/mutations
- Representative sampling of patients
- Biased selection, subsets by design
- Diagnostic “truth”
- “Soft” reference diagnosis, verification bias
- Follow-up/outcome
- Time and cost, endpoints

A participant asked about the way FDA views clinical utility in a regulatory context versus the way other stakeholders view clinical utility. Becker answered that the metrics of clinical utility used by payers or users are not necessarily needed for regulatory review of most medical devices. Clinical validity is a sufficient bar for regulatory clearance, and FDA does not make an explicit decision about clinical utility per se. Rather, it is factored into the discussions concerning safety and effectiveness. However, under the circumstance of considering claims for a device that directs the use of a specific therapy (i.e., the clinical performance of the device is tied to the performance of the drug), that is “tantamount to an aspect of clinical utility” which can be factored into consideration since that kind of trial will have demonstrated the clinical utility of the drug.

Becker highlighted some of the challenges associated with establishing analytical and clinical validity of diagnostic devices (Box 2-2). For many of the types of devices FDA has reviewed over the past 30 years, there are strong reference methods or standards available for analytical validation. This is not the case, however, for many of the emerging technologies that the agency is reviewing now, such as gene expression assays. New multivariate tests involve numerous analytes for which a full-scale analytical validation of each single analyte would be onerous and would not necessarily

be informative regarding how the test will perform as a whole. There are multiple issues regarding the clinical samples needed for validation (e.g., the types and availability of samples and their age and storage), and assessing the full spectrum of a device's performance is also a challenge. Clinical validity can present similar challenges regarding the availability of samples or patients, with issues potentially arising from sample and verification bias. There may also be limits to the "diagnostic truth" of the test results if information is limited on the patient's underlying condition, and follow-up and outcome studies may be difficult and costly to perform.

Diagnostic tests may be submitted to FDA as stand-alone tests or in association with a drug. Becker described a supplemental application for a new use for the already-marketed drug, Herceptin, where the sponsor was seeking a label indication for gastric carcinoma. Included in the application was significant additional analytical information to validate a diagnostic test for stomach cancer. This information was factored into the clinical evaluation of the drug in a collaborative review by CDRH and the FDA Center for Drug Evaluation and Research.¹

Fostering Progress in Analytical and Clinical Validation

Analytical validation of new diagnostic devices would be facilitated by standards and a technical assessment of the technology behind the tests. Becker cited the FDA MicroArray Quality Control (MAQC) project as one example of efforts in this area. In the case of MAQC, the National Institute of Standards and Technology, the National Institutes of Health (NIH), industry partners, and others are working with FDA to develop standards and quality measures for microarrays. There is also a need for better coordination of the analytical and clinical specifications during test design, verification, and validation, especially with regard to medical decision points (i.e., the test's analytical performance needs to align with the clinical performance). Becker also mentioned ongoing initiatives by government, industry, patient groups, and others that are addressing various analytical sample-related issues, such as collection, storage, and annotation.

In the quest for clean and cost-effective clinical validation of diagnostic devices, there is potential for improvement in appropriate clinical sample specification, acquisition, retention, maintenance, and accessibility. For example, clinical validation can be achieved more effectively by having well-matched intended use, claims, and evidence and choosing an intended use that is more amenable to studies that will directly support that intended use. The FDA is also exploring how to promote better coordination

¹ See http://www.accessdata.fda.gov/drugsatfda_docs/applletter/2010/103792s5250ltr.pdf for details on approval.

dination among the therapeutic product developers, the diagnostic device developers, and reviewers through pre-submission meetings or through collaborative reviews. Finally, the use of more efficient study and trial designs (e.g., adaptive or Bayesian designs) can also help facilitate clinical validation.

PAYER PERSPECTIVE

Margaret Piper of the Blue Cross and Blue Shield Association Technology Evaluation Center (TEC) described the evaluation criteria that TEC uses for its assessment of genetic test evidence. Established in 1985, TEC is housed within the Blue Cross and Blue Shield Association, which is the membership organization for Blue Cross and Blue Shield Plans. The mission of TEC is “to provide health-care decision makers with timely, objective, and scientifically rigorous assessments that synthesize the available evidence on the diagnosis, treatment, management, and prevention of disease.” TEC does not work directly for the Plans, but the Plans help to identify topics of interest for TEC to work on. TEC also makes all of its assessment products publicly available.

Technology Evaluation Criteria

Every TEC assessment involves evaluating the technology according to five general criteria,² Piper explained:

- The technology must have the required final approval from the appropriate government regulatory bodies. (This includes FDA clearance as required or CLIA certification as it applies to laboratory-developed tests.)
- The scientific evidence must permit conclusions concerning the effect of the technology on health outcomes.
- The technology must improve the net health outcome.
- The technology must be as beneficial as any established alternatives.
- The improvement must be attainable outside the investigational setting.

To evaluate genetic testing, TEC uses the ACCE framework developed by the Centers for Disease Control and Prevention (CDC) National Office of Public Health Genomics.³ ACCE refers to the four components of the

² See <http://www.bcbs.com/blueresources/tec/tec-criteria.html> for details on TEC criteria.

³ See <http://www.cdc.gov/genomics/gtesting/ACCE/index.htm> for further details on the ACCE framework.

framework—analytic validity; clinical validity; clinical utility; and ethical, legal, and social implications—which are addressed with a set of targeted questions. All TEC assessments are also reviewed by an outside medical advisory panel which has final say over the conclusions.

A Focus on Outcomes

Piper clarified the distinction that TEC makes between clinical validity and clinical utility. Clinical validity is the *association* of a test result with an outcome (e.g., diagnosing disease or predicting drug response) and is described by measures of association, such as sensitivity, specificity, and predictive value as well as odds ratios, risk ratios, and logistic regression analyses, that describe whether a genetic test retains significance when analyzed along with other criteria. Clinical validity is concerned with the significance of the test for *populations* of patients. Measures of association only quantify how well the test discriminates between populations with and without the selected outcome, and these measures alone are not sufficient to gauge the clinical usefulness of a test. Piper cited a study showing the limitations of the odds ratio in gauging test performance as an example of why single measures of association are not sufficient (Pepe et al., 2004).

In contrast, clinical utility describes the *impact* of using the test on patient management and outcomes compared to usual care and therefore describes the significance of the test for *individual* patient decision-making. From a payer perspective, the focus is on clinical utility. TEC seeks evidence that a genomic test can be used for individual patient management, and it assesses the incremental value of adding the test to usual clinical practice (measured in terms of outcomes). The ideal approach is to obtain direct evidence through randomized controlled trials (outcomes using the test versus not using the test), but such trials are seldom possible. More often, the approach is to establish an indirect evidence chain, such as the assessments conducted by the Evaluation of Genomic Applications in Practice and Prevention (EGAPP) initiative.⁴

Analytic validity describes the technical performance of the test—the accuracy of the test in measuring the analyte of interest, test repeatability, and reliability in the presence of interfering substances as well as over time. Analytic validity is carefully evaluated by the FDA when tests are submitted for marketing clearance, but such information is not routinely available for laboratory developed tests. “Most laboratory developed tests do not publish their analytic validity, do not make it otherwise publicly available,

⁴ EGAPP is discussed further by Calonge below. See also <http://www.egappreviews.org/workingrp/reports.htm> for more information on EGAPP methodology and reports. Constructing chains of evidence is discussed further by Ransohoff in Chapter 3.

and therefore we don't have any evidence of the long-term reliability of the test," Piper said.

A test should meaningfully improve discriminatory ability when added to existing predictors, or, if it is replacing a currently used test, it should demonstrate superior discrimination. This is easiest to evaluate when test results are classified in a manner that informs decision making (high risk versus low risk). One statistical approach to evaluate the discriminatory ability of a test is concordance, or the "*c*-statistic," which is the area under the receiver operating characteristic curve (a measurement that compares the sensitivity of a test to its false positive rate as the discrimination threshold is altered). This is not a very powerful method of analysis, Piper noted, and it can be difficult to determine if improvement in the *c*-statistic is clinically meaningful with regard to treatment decisions. Another method is the use of a reclassification analysis (Pencina et al., 2008). Risk is first classified by standard methods and then reclassified with the additional information provided by the genetic test results incorporated into the analysis. The net reclassification improvement is then calculated. Analysis must take into account whether the risk is reclassified correctly or incorrectly.

Case Examples of TEC Assessments

Oncotype DX Assay

The Oncotype DX assay is used for predicting response to chemotherapy in women with node-negative, estrogen receptor (ER)-positive breast cancer. TEC first evaluated the relationship between the Oncotype DX Recurrence Score and distant disease recurrence within 10 years. A study published in 2004 first established clinical validity (Paik et al., 2004a), Piper said, but the evidence was deemed insufficient to meet the TEC criteria (TEC, 2005). A subsequent study published in 2006 established the relationship between the Recurrence Score and the likelihood of benefit from chemotherapy (Paik et al., 2006). Again, this evidence was primarily supportive of clinical validity and was not sufficient to meet TEC criteria, Piper said, but the TEC medical advisory panel asked if a better analysis had been done on the existing data.

Reclassification analyses presented in a poster and partially published in a review article found that about half of the patients who were originally classified by the National Comprehensive Cancer Network (NCCN) criteria as being at a high risk of recurrence were subsequently reclassified as being at low risk of recurrence by Oncotype DX testing (Table 2-1) (Paik et al., 2004b).

Before Oncotype DX testing, Piper explained, all of these low-risk patients who were classified as high risk by the original NCCN criteria

TABLE 2-1 Oncotype DX Reclassification of Patients

Classification by NCCN	Reclassification by Oncotype DX	n	% Distant Recurrence Free at 10 years (95% CI)
Low (8%)	Low	38	100 (NR)
	Intermediate	12	80 (59–100)
	High	3	56 (13–100)
High (92%)	Low	301	93 (89–96)
	Intermediate	137	86 (80–92)
	High	178	70 (62–77)

SOURCE: Adapted from TEC, 2008a.

would have received chemotherapy. Since the reclassification analysis was done using retrospective data from completed clinical trials, outcomes were known. The confidence interval for no recurrence in the reclassified population was 89 to 96 percent. The absolute benefit from anthracycline chemotherapy in these low risk patients is 1 to 3 percent at best, she said. In addition, the lower the prior risk of recurrence as indicated by the Oncotype DX result, the less absolute benefit the patient derived from chemotherapy. At such low absolute chemotherapy benefit, the harms may be perceived as greater than the benefit and a woman might reasonably choose to avoid chemotherapy. This analysis, together with the prior information, allowed the test to meet the TEC criteria (TEC, 2008a).

Genetic Test for Long QT Syndrome

Another example of a TEC assessment concerns congenital long QT syndrome, which can lead to major cardiac events and sudden death. Although beta-blockers are effective as a low-risk preventative treatment, the clinical diagnostic criteria are not well established, and the syndrome is difficult to detect. There is, however, a genetic test that can detect 60 to 70 percent of people with long QT syndrome. Individuals who are defined as having long QT syndrome by genetic marker testing only (i.e., no clinical signs) also have a high risk of catastrophic cardiac events (Moss et al., 2007). If an individual has been diagnosed with congenital long QT syndrome, either following an event or by clinical means, or if there is a known mutation in the family, then relatives with possible long QT syndrome can be identified through genetic testing for the mutation and treated with beta-blockers to reduce the risk of adverse cardiovascular outcomes (Roden, 2008). The disease can be ruled out with confidence for those who test negative for a known mutation. Based on these findings, the TEC assessment was affirmative for use of the genetic test (TEC, 2008b).

Epidermal Growth Factor Receptor Mutation Testing

As a final example, Piper described TEC's assessment of whether epidermal growth factor receptor (EGFR) mutations can prospectively predict response to tyrosine kinase inhibitor therapy for patients with non-small-cell lung cancer. The majority of the early data relate to the drugs gefitinib (which is no longer available in the United States) and erlotinib, both of which inhibit EGFR activation. The first assessment of genetic testing to predict response to EGFR inhibitors did not meet TEC criteria because there were not enough data to separate the responder and nonresponder populations through mutation testing (TEC, 2007). It also appeared that the test did not reliably identify nonresponders, as some patients with wild-type EGFR genes who were not supposed to respond to therapy actually did respond (TEC, 2007). There was also a concern that the inhibition mechanisms of gefitinib and erlotinib might be slightly different. As only erlotinib is available in the United States, the TEC medical advisory panel requested independent assessment of erlotinib, and the subsequent assessment concluded that EGFR mutation testing to predict response to erlotinib treatment does meet TEC criteria. Outcomes for progression-free survival and overall survival showed a much better separation between responders and nonresponders, and showed that patients with the wild-type EGFR gene were not likely to respond to erlotinib, thus, indicating it is best for them to move to an alternative treatment (TEC, 2011).

EVIDENCE-BASED REVIEW GROUP PERSPECTIVE

The EGAPP initiative is an independent, non-regulatory CDC-funded project to develop evidence-based recommendations on the appropriate use of genetic tests, said Ned Calonge, chair of the EGAPP Working Group. The EGAPP process is transparent and publicly accountable, and its methods integrate knowledge from existing processes of evaluation and appraisal, such as those of the ACCE systematic review process, the Evidence-Based Practice Centers program of the Agency for Healthcare Research and Quality (AHRQ), and the U.S. Preventive Services Task Force, to assess the quality of individual studies, the adequacy of overall evidence, and the magnitude of net benefit. EGAPP also takes contextual issues into account. In addition to providing recommendations, EGAPP identifies gaps in the evidence in order to inform the research agenda.

The EGAPP Working Group Process

The basic steps in the EGAPP working group process are:

- Select the topic or genomic application for evaluation.
- Define the clinical scenario (diagnosis, disease screening, risk assessment, prognosis, or pharmacogenetics).
- Create an analytic framework of key questions to guide the evidence review.
- Find, synthesize, and evaluate the quality and adequacy of existing literature.
- Determine net benefit (benefit minus harms) of test application.
- Create a recommendation based on the certainty of net benefit.

Analytic Framework

Often, Calonge noted, there is no overarching or direct evidence of health outcomes associated with the use of a genomic test (such as the evidence from a randomized controlled trial). In these cases, EGAPP uses the ACCE criteria to develop an indirect chain of evidence to address key questions of analytic validity, clinical validity, and clinical utility. As an example, Calonge presented an analytic framework and a set of key questions for the assessment of pharmacogenomic testing for selective serotonin reuptake inhibitor (SSRI) therapy (Figure 2-1) (Teutsch et al., 2009).

Quality of Evidence

After developing the framework and the key questions, the working group grades the quality of the evidence. The quality assessment takes into account the hierarchical level of the study design as well as study flaws and threats to internal validity, Calonge said. Evidence is classified as either *convincing* (the observed effect is likely to be real), *adequate* (a higher risk exists that the effect may be influenced by study flaws), or *inadequate* (too many flaws exist to confidently assign the results to the factors under study).

Net Benefit

Determining the net benefit involves balancing the benefits or potential benefits with the harms or potential harms. This often requires comparing harms and benefits that are very different in terms of health, value, or metrics, Calonge said. Net benefit is classified as *small*, *moderate*, or *substantial*. Based on the overall assessment of evidence, EGAPP also determines the certainty of net benefit. Calonge described certainty as the opposite of the risk of being wrong. The higher the level of certainty, the less likely that the recommendation is incorrect or will be changed because of future research. An example of high certainty is the value of blood pressure screen-

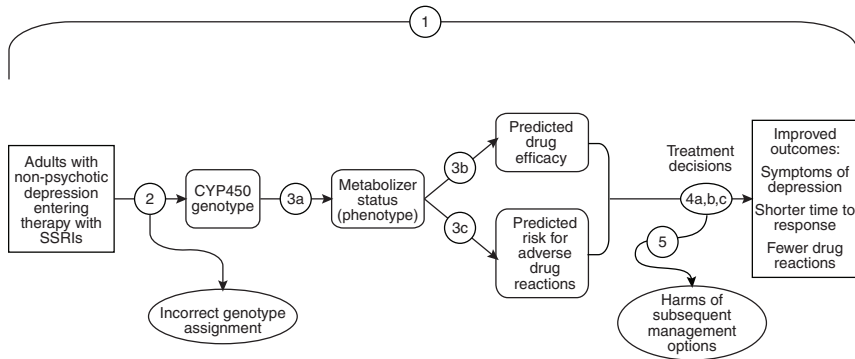


FIGURE 2-1 Sample analytic framework.

NOTE: Key questions (shown as numbered circles) addressed at each stage are: (1) overarching, direct evidence; (2) analytic validity [technical performance, including analytic sensitivity/specificity, reliability, and assay robustness]; (3) clinical validity [ability to identify or predict the disorder of interest: clinical sensitivity, specificity, and predictive value relating to expression/phenotype]; and (4 and 5) clinical utility [balance of benefits and harms with clinical use: efficacy/effectiveness and net benefit].

SOURCE: Teutsch et al., 2009.

ing. It is highly unlikely that at some point in the future it will be decided that people should not be screened and treated for high blood pressure. Moderate certainty means that there are some questions about the evidence and some risk that future research could lead to a change in the recommendation, but in the judgment of the EGAPP working group the evaluation of net benefit has met the criteria for making the recommendation. Finally, low certainty is when there is inadequate evidence of a net benefit to make a recommendation.

A Focus on Net Benefit

If there is high or moderate certainty concerning a small to substantial net benefit, EGAPP will recommend use of the genomic application, Calonge said. If there is high or moderate certainty of a zero benefit, or of a net harm, EGAPP recommends against the use of the application. Low certainty results in a conclusion of insufficient evidence and no recommendation for or against is made.

EGAPP further classifies insufficient evidence conclusions according to contextual factors. Such applications are classified as *neutral*, indicating that it is not possible to predict what future research will find; *discouraging*,

in cases where either the risk of harm is so high that EGAPP discourages use of the application until specific knowledge gaps are filled or the topic is not likely to ever meet evidentiary standards; or *encouraging*, in cases where the working group believes that the test is likely to meet evidentiary standards with further study and reasonable use in limited situations is appropriate while awaiting further evidence.

Case Examples of EGAPP Assessments

Lynch Syndrome

Lynch syndrome, also known as hereditary nonpolyposis colorectal cancer, is an inherited condition that increases an individual's risk of colon and other cancers, including those of the stomach, small intestine, liver, gallbladder ducts, upper urinary tract, brain, skin, prostate, endometrium, and ovaries (IOM, 2010). The EGAPP working group “found sufficient evidence to recommend offering genetic testing for Lynch syndrome to individuals with newly diagnosed colorectal cancer in order to reduce mortality and morbidity in relatives.” This is a very specific use, Calonge noted.

Three genetic testing strategies showed evidence of analytic validity: microsatellite instability (MSI) testing, immunohistochemistry (IHC) testing, and testing for mutations in the BRAF gene. The analytic performance of MSI testing is high, according to College of American Pathologists (CAP) external proficiency testing. IHC testing for Lynch syndrome mismatch repair gene proteins is not currently subject to CAP testing, but IHC testing for other proteins is, and therefore an assumption was made that IHC testing for Lynch syndrome has adequate analytic validity. Finally, BRAF mutation testing is based on single-gene mutation sequencing, and analytic validity in that setting is high. Based on existing clinical studies, EGAPP also found adequate evidence of clinical validity for all three tests.

Clinical utility was first assessed for the probands themselves, but EGAPP found insufficient evidence to support differential treatment options based on a proband being identified as having Lynch syndrome. The working group did note, however, that a small body of evidence suggests that MSI-high tumors might be resistant to treatment with 5-fluorouracil and more sensitive to irinotecan (Palomaki et al., 2009), and it highlighted this as a research gap worthy of further study.

Lynch syndrome testing had the greatest clinical utility, Calonge said, with regard to first- and second-degree relatives of patients testing positive for Lynch syndrome. Seven studies found an increase in colonoscopy testing of between 53 and 100 percent in test-positive relatives (Palomaki et al., 2009). The benefit outweighed the harms, which were judged as no more harmful than colonoscopy testing in general. Evidence for utility was pro-

vided by a randomized controlled trial that showed a 62 percent reduction in colorectal cancer and related mortality in relatives with Lynch syndrome mutations who then chose colonic surveillance (Jarvinen et al., 2000) and from an observational study showing a 73 percent mortality reduction in a study involving nearly 3,000 persons across 146 Lynch syndrome families (de Jong et al., 2006). There was similar indirect evidence for increased ovarian and endometrial cancer screening in women (Schmeler et al., 2006). Together, these pieces supported the EGAPP recommendation for Lynch syndrome testing (Palomaki et al., 2009).

Breast Cancer Tumor Gene Expression Profiles

Calonge also described the EGAPP assessment of breast cancer tumor gene expression profiles. In this case, after a review of the Oncotype DX, MammaPrint, and H:I ratio tests, EGAPP “found insufficient evidence to make a recommendation for or against the use of tumor gene expression profiles to improve outcomes in defined populations of women with breast cancer.” This insufficient evidence conclusion was classified as encouraging, however, as EGAPP “found preliminary evidence of a potential benefit of testing results to some women who face the decisions about treatment options (reduced adverse events due to low risk women avoiding chemotherapy) but could not rule out the potential for harm for others (breast cancer recurrence that could have been prevented).”

The evidence regarding analytic validity was deemed inadequate. There were some data on technical performance, but it was not possible to make estimates of analytic sensitivity or specificity, and testing failed on 14.5 to 19 percent of fresh samples. Despite the inability to establish analytic validity, there was enough evidence for the clinical validity of Oncotype DX based on three studies and for the clinical validity of MammaPrint based on two studies (Marchionni et al., 2008).

There were no studies of clinical utility for MammaPrint, which was a critical evidence gap, Calonge said. In contrast to the conclusion by TEC that Oncotype Dx met their evaluation criteria, the EGAPP assessment of a retrospective analysis of one arm of a prospective clinical trial found the evidence promising but ultimately unconvincing. And although women are likely to benefit by avoiding unnecessary chemotherapy, the potential for harm (recurrence and perhaps death) is significant for a small number of low- and intermediate-risk women who might benefit from chemotherapy but forego it based on test results. There were no data available on use in women with high risk on conventional assessment but with low risk on Oncotype DX (the type of reclassification analysis evaluated by TEC).

The conclusions were that there was encouraging indirect evidence for Oncotype DX (Marchionni et al., 2008) and a plausible potential use for

MammaPrint (Marchionni et al., 2008). Calonge noted that two randomized controlled trials are under way, the results of which may allow for updating the recommendation.

Common Research Gaps

In closing, Calonge listed several hurdles to evaluation. One challenge to establishing analytic validity is that necessary information is often missing or unavailable due to proprietary issues or because these may be laboratory-developed tests. Clinical validity is often based on testing subjects with the potential for sources of bias (e.g., selection process, study design) to influence the results. The clinical utility evaluation is a major source of insufficient-evidence conclusions. Furthermore, there are few randomized controlled trials, observational studies present bias issues, and recommendations based on observational study results run a higher risk of being wrong than those based on randomized controlled trial findings.

HEALTH-CARE PROVIDER PERSPECTIVE

A Focus on Value

As an oncologist, Daniel Hayes of the University of Michigan Comprehensive Cancer Center looks at genomic diagnostic tests from the perspective of value. The goal is to improve cancer outcomes by focusing the “right therapy on the right patient,” thereby increasing the chances of cure, survival, or palliation and decreasing exposure to toxicity from unnecessary or inappropriate therapy.

Hayes began with data showing that the odds of dying of breast cancer in the developed world gradually increased from 1950 until about 1980, reached a plateau for about a decade, and then began a steady decline around 1990 (Peto et al., 2000) (see Figure 2-2). Hayes said that the plateau and early decline was due not to screening, which did not begin until the late 1980s, but rather to the widespread application of adjuvant chemotherapy and hormone therapy which began in the late 1970s and early 1980s. Subsequent analysis has shown that the overall decline can be attributed equally to the use of screening and the use of adjuvant therapy (Berry et al., 2005) (see Figure 2-2).

The decision to treat differs from patient to patient and is affected by patient, provider, and societal perspectives regarding risks, benefits, and costs. There is no clearly defined level of benefit below which treatment becomes “not worth it.” Two patients presented with the same Oncotype DX Recurrence Score, for example, may come to different treatment decisions.

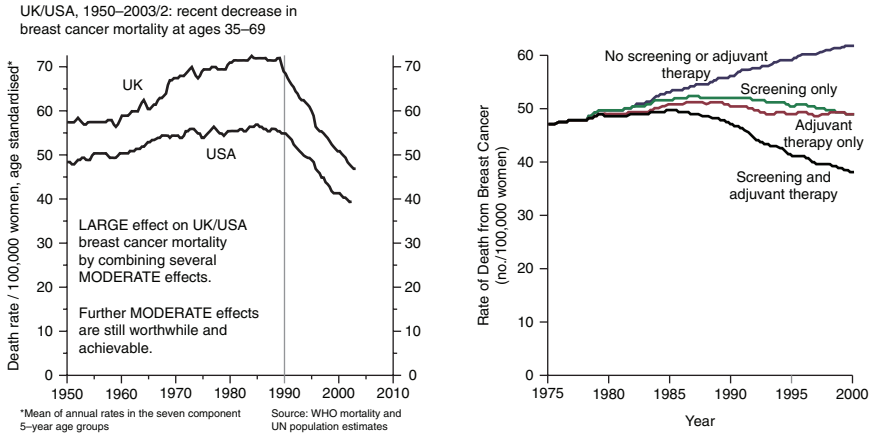


FIGURE 2-2 Breast cancer mortality rates from 1950-2003 and the effect of screening and adjuvant therapy.

SOURCE: Berry et al., 2005; Peto et al., 2000.

To help guide decisions regarding the value of genomic diagnostic tests, the American Society of Clinical Oncology (ASCO) developed tumor marker guidelines (Harris et al., 2007). The guidelines are rather conservative, Hayes said, and the ASCO panel recommended only those markers for which results would change clinical decisions. To facilitate evidence analysis, the panel developed a tumor marker utility grading scale which classifies studies into one of five levels of evidence (LOEs). LOE I studies are prospective, highly powered studies designed specifically to test a tumor marker or a meta-analysis of lower-level studies. LOE II studies are prospective studies of a therapeutic product where the study of the marker is a secondary objective. LOE III through V are retrospective analyses. Only level I or level II studies should be used for evidence evaluation. Unfortunately, Hayes said, most of the tumor marker studies that are available are LOE III.

For those already diagnosed with cancer, a tumor marker is clinically useful when it is prognostic or predictive of cancer outcomes or predicts toxicity; when the magnitude of the effect is sufficiently large that clinical decisions based on the test data result in acceptable outcomes; and when the estimate of that magnitude of effect is reliable. Cancer diagnostics allow clinicians to identify those patients for whom the benefits of a treatment do not outweigh risks, in which case they can safely recommend withholding that treatment. “We are trying to identify patients who would forgo or discontinue therapy to avoid toxicities,” Hayes said. As an example, he noted that ER testing of tumors is regularly done in the clinic because tamoxifen is effective only in ER-positive patients (Berry et al., 2006), which allows

ER-negative patients to avoid the toxicity associated with treatment. However, only half of ER-positive patients are responsive to the therapy, and the test does not indicate which particular ER-positive patients will benefit from treatment.

Evidence Generation

In oncology, evidence is generated in two ways, Hayes said. The first is by developing prospective clinical trials to “test the test,” where the marker is the primary objective. There are very few such trials to test cancer diagnostics, however, as they are large, costly, and lengthy. Another, more innovative way to generate evidence is by using archived specimens.⁵ Hayes said that ideally these specimens would be from a previously conducted prospective trial and when testing predictive factors, the specimens would have originated from a trial that specifically addressed the effectiveness of the therapy in question in a randomized fashion. Such studies, when designed and conducted with as much rigor as one would put into a prospective trial and when appropriately confirmed, can generate LOE I data. Hayes and colleagues have authored a proposal to revise the LOE scale to provide a more detailed account of the use of archived specimens to generate a sufficiently high level of evidence to achieve clinical utility (Simon et al., 2009).

One of the barriers to generating the necessary evidence is that tumor marker research, especially the clinical component, is not perceived to be as exciting or as important as research on new therapeutics. Marker studies garner less academic credit and less funding and are often conducted with less rigor. There is also less evidence required for clinical use (e.g., by FDA or guideline panels), less quality control/quality assurance and proficiency testing, and much less reimbursement.

This has led to a vicious cycle, Hayes said. Tumor marker utility is poorly valued, which leads to the low level of reimbursement. This means lower funding for tumor marker research and little incentive to do properly designed and controlled clinical studies. The lack of trials leads to lower levels of evidence and less certainty concerning the data, resulting in few recommendations for use, and the cycle repeats.

Breaking the cycle of undervalue, Hayes said, requires increasing the level of reimbursement for marker testing, increasing funding for tumor marker research, and creating incentives to conduct well-designed trials. This will result in the generation of level I evidence that can support guidelines and recommendations for use, which in turn will lead to marker utility that is highly valued. Hayes called for tumor marker publications to have increased rigor, comparable to that of therapeutic trials. He also suggested

⁵ Discussed further by Simon in Chapter 3.

reforms to the regulatory review of tumor markers, including requiring analytical validity and clinical utility, eliminating laboratory-developed test discretion, and requiring that new drug registration trials include a bio-specimen bank. Overall, he concluded, it is necessary to think about tumor markers in the same way that therapeutics are considered.

PANEL DISCUSSION

Applying Population-Based Data to Individuals

Participants discussed the challenges of applying results from clinical studies and population-based data to individual patients. When applying new risk stratification methodology, there will inevitably be some patients who were correctly classified with the old methodology but who are incorrectly classified with the new methodology. “How do you deal with patients who actually are disadvantaged as we move to what is, from a population standpoint, a better test?” asked a participant. Hayes said that the focus should be to improve the odds of being cured. He referred participants to *Adjuvant! Online* (<http://www.adjuvantonline.com>) which allows providers to enter patient data and make estimates of various risks (e.g., odds of recurrence, benefit from systemic adjuvant therapies, and adverse events). Quantifying those risks allows the patient to better understand the treatment options being offered, Hayes said. Calonge added that while the goal is to balance the potential benefits and harms, people tend to be benefit oriented and do not examine harms as often or in as much depth as they examine benefits. The potential harms include providing therapy to someone who does not need it. In effect, “the cost of the benefits for a person who would have benefited from therapy in the old scheme is the harms to all those who did not benefit.” Becker noted that survival curves are not step functions and that some patients in the apparently responsive group do die. It is always difficult to project what a population study means at the individual patient level. Another challenge to consider is how to move forward when tests results suggest actions in different directions (e.g., test A predicts a high risk, while test B indicates a low risk).

Clinical Utility

Another topic of discussion was clinical utility versus personal utility, which refers to the value of information to the individual for use in decision making. Calonge noted that the potential value of information is included in the EGAPP outcomes set, but understanding the actual value is difficult. If ending the diagnostic odyssey provides a health benefit to an individual,

what is the value of that benefit, and how much should we pay for that additional information?

Hayes added that many patients have tests done primarily because the tests are available to them, under the presumption that they will predict something and that there will be an action taken on the results. In some cases, there is a marker and an implied outcome, but there is not a high level of evidence that the outcome is associated with the marker or that some action should be taken based on the marker. Hayes said that patients are often mistreated based on presumed information, and many people make decisions that end up harming them instead of helping them. The role of a physician is to help patients avoid making decisions that will harm them. From a societal perspective, it is not prudent to spend a lot of money on tests that, while new and exciting, are still unproven. Tests should be done because they have clinical utility. One could argue that if the patient is willing to pay for the test, then it is the patient's right to have it done. But issues appear downstream with the potential for mistreatment and with the added costs to the system of such treatment (e.g., to payers and to the insured in the form of increased premiums). A participant noted that genetic counseling is based on the concept of personal utility. Genetic counselors spend a lot of time talking about the patient's motivation for wanting the test. Performing the procedure may not change the result of a particular situation, but people may "perceive what they want to perceive" even if there is not a direct health outcome.

Piper added that insurance companies look at health outcomes and draw fine lines for what qualifies for reimbursement. Information for information's sake is not health outcome oriented. Similarly, information that is of value for life planning (e.g., making wills, buying insurance, or making job decisions) is not health outcome oriented. A health outcome would instead be a case when the information reduces the diagnostic odyssey or is useful for reproductive decision making. However, if the information predicts the risk of a disease for which there is no preventive or ameliorative treatment, then there is no health outcome. "Where personal preferences come in, is in the final decision-making after you have all the data," said Piper.

Evidence

Participants also discussed whether there is enough commonality across stakeholders to set a common evidentiary bar. In general, panelists felt that the different stakeholders who were represented were not that far apart. Calonge noted that EGAPP has started looking at the reclassification approach described by Piper, but it has been very careful not to drop the evidentiary standard too low and thus increase the risk of being wrong and

harming individuals. Bringing critical appraisal questions concerning coherence, consistency, strength of association, and precision of the estimate into observational or post hoc analysis studies will allow evidence to be moved closer toward adequate and convincing levels.

Piper said that there is no one blueprint to follow in all situations. The analysis for long QT syndrome was different from the analysis for Onco-type DX, for example. The analysis must be adapted to the clinical setting and the evidence needed for that particular application. The bottom line is to establish that there is evidence of improved patient outcomes. One does not stop at clinical validity and associational evidence, Piper said.

Becker agreed that there is a fair amount of commonality about the kind of evidence that helps in reaching a decision. One of the things that sets the FDA approach apart from the approaches that can be employed in other settings, he said, is that the agency needs to look at tests on a device-by-device, test-by-test basis and make regulatory decisions about the individual test that the device sponsor has brought to the agency. Stakeholders outside of FDA are more able to synthesize information across the literature and across tests. However, there are some circumstances in which FDA can handle class-specific issues across all of the devices of a particular type, rather than individually for each specific test.

Participants encouraged including subject matter experts in the process of evidence analysis. While such experts may not be part of the final evidence decision, they should be part of the gathering of the evidence.

3

Approaches to Evidence Generation

Key Points Raised by Speakers

- The translation of diagnostic tests into practice needs to be facilitated with evidence of effectiveness in a clinical setting.
- Studies should be designed with a specific clinical context in mind, and benefits, harms, intended use, and desired test features should be considered.
- More efficient methodologies for generating evidence need to be developed to expedite decision-making.
- Researchers need to collect and store biospecimens from prospective clinical trials with future analysis and use in mind.

PHARMACOGENOMICS CLINICAL TRIALS

A basic question in clinical research is whether an intervention works across populations. In pharmacogenomics research, said Caryn Lerman of the University of Pennsylvania, the question can be reframed as whether the intervention benefits or harms particular patients. Ultimately, the question is whether a genomics-based therapy is worth doing from the perspectives of patients, payers, and other stakeholders. Data to answer these questions can be gathered through observational studies of the association of a genomic marker with an outcome (e.g., a cohort study) or through experi-

mental studies of the efficacy of a pharmacogenomic intervention based on accumulated data (e.g., a randomized controlled trial).

Randomized Controlled Trials

An advantage of randomized controlled trials (RCTs) over cohort studies is that they provide controlled exposure to treatment. In addition, randomization helps avoid the type of confounding that can occur in an observational study, in which treatment may be selected based on patient characteristics.

A retrospective trial of a pharmacogenomic marker is carried out after a RCT of a drug has been completed, with researchers testing patient samples to identify which patients were positive or negative for a particular marker and then comparing that information with the patients' responses to the drugs being tested (Figure 3-1). Retrospective trials can provide useful

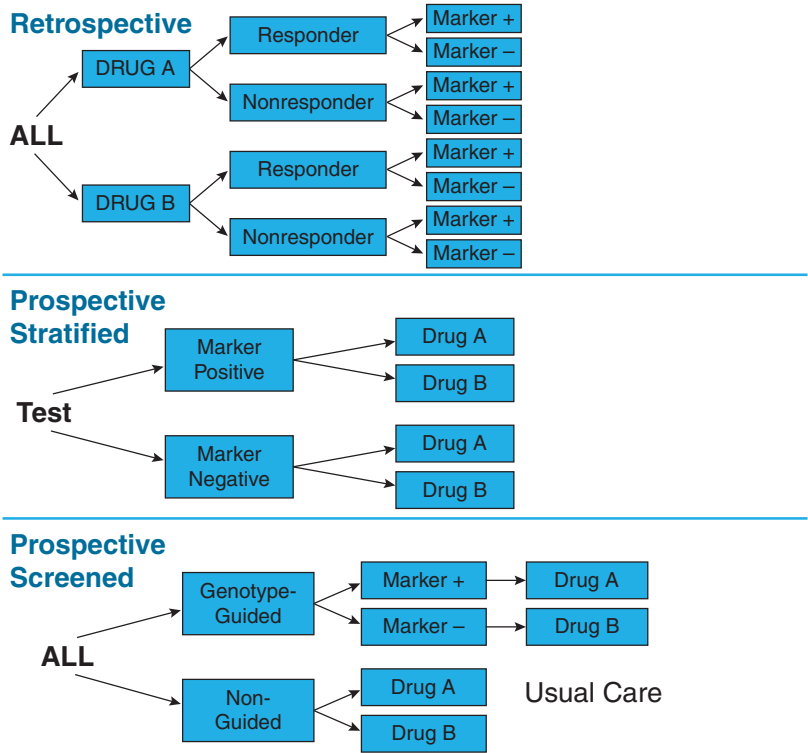


FIGURE 3-1 Pharmacogenomic trial designs, including retrospective, prospective stratified, and prospective screened.

SOURCE: Lerman, IOM workshop presentation on November 17, 2010.

data when a marker is unknown at trial initiation. They are also ideal for hypothesis generation and can be used for independent validation. There are, however, several important limitations to relying solely on retrospective clinical trials, including unbalanced groups, reduced power based on those unbalanced groups, and missing data (e.g., not all patients may have consented to tissue collection or use of their tissue for further study).

In contrast, *prospective stratified* trials first test and identify participants as marker positive or marker negative and then randomize each group of participants to therapy arms (i.e., all the marker-positive participants are randomized to either a drug group or a control group, and the same is done for the marker-negative participants) (Figure 3-1). The advantage of this design is that the trial is based on a hypothesis that takes into account prior genomic knowledge about the members of the test population. One important feature of prospective stratification is that it allows for enrichment of more rare genotype groups and balancing of treatment assignment.

A third variant is the *prospective screened* trial, which, Lerman said, comes closer to a clinical utility model than to a clinical validity model. Some argue, she said, that this is the true test of whether personalized medicine works. In a prospective screened trial, patients are randomized to either a genotype-guided group or a non-guided group. In the genotype-guided group, participants are tested for the presence or absence of the marker under study and then assigned to a treatment group based on the hypothesized association of the marker with the outcome of a particular therapy. The therapy for those who are marker-negative can be an alternate therapy. Results for the genotype-guided groups are compared with those for the non-guided group, which is either randomized to the same two therapies as the guided group or receives the current standard of care (Figure 3-1). The prospective screened design has high ecological (i.e., real life) validity, providing evidence of whether a genome-guided therapy will provide significantly better outcomes than non-guided therapy.

Genome-therapeutic response associations, however, are not necessarily translated into clinical practice. To foster the adoption of genomics-based interventions, it will be important to increase the generalizability of clinical trial designs and results to include clinical practice settings; to demonstrate improvement in health outcomes as well as the cost effectiveness of testing versus not testing; and to establish evidence-based guidelines. Lerman offered several reasons for the reduced generalizability and lack of translation of classic randomized clinical efficacy trials into clinical practice. RCTs have strict eligibility criteria and are conducted in a highly controlled setting, the treatment is protocol-driven, and treatment compliance is very closely monitored. In contrast, in everyday clinical practice the population is very diverse, the practice settings are heterogeneous, treatment is flexible and depends on clinical judgment, and compliance is variable (and likely lower than in the clinical trial setting).

TABLE 3-1 Classic Randomized Controlled Trials (RCTs) Versus Practical Clinical Trials (PCTs)

	Classic RCT/Efficacy	PCT/Effectiveness
Research Question	Does it work in ideal circumstances?	Does it work under best practice conditions?
Population	Selective, homogeneous	Diverse, heterogeneous
Setting	Specialized, controlled	Clinical practice
Intervention	Fixed, protocol-driven	Flexible, clinician judgment
Comparator	Placebo or active	Usual care, least \$
Compliance	Closely monitored, high	Highly variable
Assessments	Elaborate, complex	Simple outcomes
Goal	FDA approval	Adoption in practice

SOURCE: Lerman, IOM workshop presentation on November 17, 2010.

One approach to addressing these issues, Lerman suggested, is a *practical clinical trial* model (also called a pragmatic clinical trial; [Brass, 2010]). Classic RCTs focus on establishing the efficacy of the intervention, while practical clinical trials study the effectiveness of the intervention, looking at simple outcomes, such as health outcomes, patient satisfaction, and costs (Table 3-1).

The advantages of the practical approach to clinical trials are that they are more reflective of patients and practice, more efficient and less burdensome, and the results of the trial are more likely to be generalizable. The disadvantages are that practical clinical trials are less experimentally rigorous by design, usual care is not a stable comparator, and increased heterogeneity results in a much lower signal-to-noise ratio, making greater sample sizes necessary.

Case Example

To illustrate these issues, Lerman offered a case example involving the pharmacogenetics of nicotine addiction treatment. The six-month quit rate across a variety of interventions (lozenge, gum, patch, inhaler, nasal spray, bupropion, and varenicline) is very low. Even using best-in-class pharmacotherapy with varenicline (CHANTIX[®]), only about one-third of smokers will have successfully quit smoking at 6 months (Gonzales et al., 2006).

A marker that could predict which intervention would be optimal for a given patient could have a substantial medical and public health impact, and Lerman and colleagues have validated a novel metabolic biomarker across several clinical trials. The ratio of the nicotine metabolites 3-hydroxycotinine and cotinine is a stable measure of an individual's nicotine metabolism rate derived from smoking (Ray et al., 2009). This marker reflects a heritable trait and is independent of the time since the last cigarette, and the metabolites can be measured in saliva, plasma, and urine. This metabolic marker is highly correlated with the *CYP2A6* genotype (i.e., it is a phenotypic measure of a genomic trait) (Benowitz et al., 2003; Malaiyandi et al., 2006), but the test for the metabolites is less costly and easier to perform than the genomic test. As a phenotypic test, it also reflects environmental influences on nicotine clearance, as well as genetic influences beyond *CYP2A6*. Lerman and her colleagues obtained evidence of association by a retrospective analysis of four clinical trials, and they then established the clinical validity of the nicotine metabolism ratio test in a prospective stratified RCT.

Next Lerman discussed a hypothetical practical clinical trial of genotype-guided versus non-guided nicotine therapy, comparing a nicotine patch (low cost, low toxicity) to varenicline (higher cost, greater toxicity) (Figure 3-2). Participants in the genotype-guided arm would be tested for their nicotine metabolism ratio, and slow metabolizers would be treated with a nicotine patch, while fast metabolizers would receive varenicline. Participants in the non-guided arm would be randomized to patch or varenicline.

In this hypothetical scenario, about 20 percent of smokers in the population are slow metabolizers. In the genotype-guided arm this would mean that 20 percent of participants would get the patch and 80 percent would receive varenicline. In the non-guided group, however, randomizing between the two medications means that 50 percent of these participants are treated with the patch and 50 percent with varenicline; based on this randomization, half of the slow metabolizers in the non-guided group will receive the same treatment as the slow metabolizers in the guided group, and similarly for the normal metabolizers in the two groups. To have sufficient statistical power to examine the marginal quit rates in the genotype-guided versus non-guided groups, the study would need to enroll thousands of people.

It is much more efficient to assess genomics-guided versus non-guided therapy in a prospective stratified trial, which allows for oversampling of slow metabolizers in order to achieve comparable numbers of slow and fast metabolizers in the various treatment arms (Figure 3-3). Examining efficacy is then a matter of simply comparing patch to varenicline for the slow metabolizers, and patch to varenicline for the fast metabolizers.

For illustration: n = 100/group, SM (20%) NM (80%)

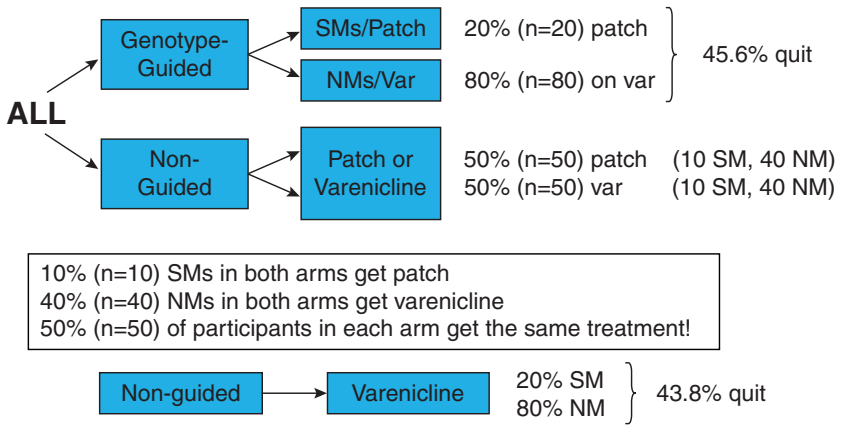


FIGURE 3-2 Hypothetical genotype-guided versus usual care scenario to measure smoking cessation rates using varenicline or patch. Abbreviations: NM (normal metabolizer); SM (slow metabolizer; var (varenicline). SOURCE: Lerman, IOM workshop presentation on November 17, 2010.

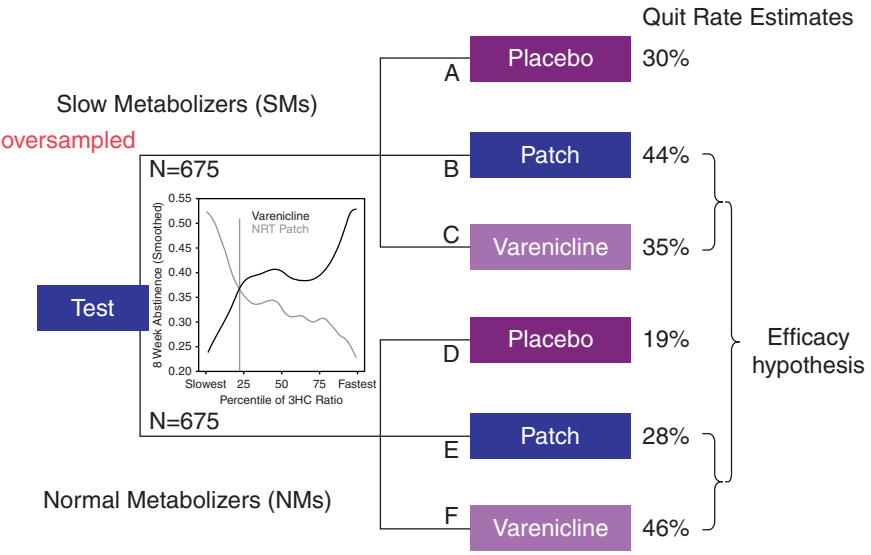


FIGURE 3-3 Prospective stratified RCT scheme. SOURCE: Lerman, IOM workshop presentation on November 17, 2010.

Principles

In summary, Lerman said that one trial design does not fit all. Retrospective and prospective RCTs are both valid, but prospective trials overcome the limitations of retrospective trials, and population enrichment designs can be used.

Practical clinical trials are an important tool to address the translational gap. However genotype-guided versus non-guided trial designs are inefficient under some scenarios and are not likely to supplant classic RCTs. Once clinical validity has been established in a prospective trial, a practical clinical trial could be part of the validation pathway to help facilitate the transition into practice.

USE OF ARCHIVED SAMPLES TO EVALUATE GENOMIC TESTS

Richard Simon of the National Cancer Institute (NCI) described his work with colleagues on the use of archived specimens to generate new evidence about the clinical utility of prognostic and predictive biomarkers (Simon et al., 2009).

Biomarkers

The term *prognostic biomarker* is not well defined, Simon noted, and it is used differently in different fields. For the purposes of pharmacogenomics, Simon suggested that prognostic biomarkers are biomarkers that are measured before treatment and are used to predict the long-term outcome for patients receiving standard treatment. The marker may provide information about both the aggressiveness of the disease and the effect of the treatment. A primary intended use of the prognostic marker is to define a subset of patients who have a very good prognosis on the standard treatment and therefore do not require a more aggressive treatment.

An example of the application of a prognostic biomarker is the Oncotype DX gene expression assay initially developed for node-negative, ER-positive women who are receiving tamoxifen. The goal of testing is to identify those patients who are unlikely to benefit from adjuvant chemotherapy in addition to surgery/radiotherapy and hormonal therapy. The Oncotype DX test was initially validated through a retrospective analysis of a previously performed prospective clinical trial. The key to the successful development of the test was that it was done with an indication in mind, Simon said. An important therapeutic decision context was identified, the development and clinical validation separated in a staged manner, and analytical validation emphasized. According to Simon, most prognostic factor studies are not performed with a specific clinical context defined prior to

starting and are consequently very difficult to interpret. A prospective trial, TAILORx, for the validation of Oncotype DX is ongoing (Zujewski and Kamin, 2008).¹

Predictive biomarkers are measured before treatment to identify who will or will not benefit from a particular treatment. Many cancer treatments benefit only a minority of the patients to whom they are administered, Simon said, and there is probably no case in which a treatment does not harm at least some of the patients. Being able to predict which patients are likely to benefit (or which are unlikely to benefit) could save patients from unnecessary toxicity, enhance the chances of success, and help control medical costs. Predictive biomarkers are also a critical part of the drug development process for almost all new cancer drugs.

Predictive biomarkers are usually single gene/single protein markers, such as with HER2 testing to determine the appropriateness of anti-HER2 breast cancer treatments (e.g., Herceptin) (Baselga et al., 1999; Wolff et al., 2007) and KRAS analysis to determine appropriate usage of anti-epidermal growth factor receptor (anti-EGFR) antibodies in treating colorectal cancer (Lee and Chu, 2007).

Validation

Validation is essentially a showing of fitness for intended use. Validation is often broken down into analytical validation, clinical validation, and clinical utility. There is some ambiguity concerning what people mean when they talk about these different terms, especially clinical utility, Simon said. Clinical utility can take into account costs or advantages and disadvantages, but he said that the key factor in utility is whether the result of the test is actionable and informs treatment selection to the benefit of the patient.

The optimal designs for evaluating the clinical utility of a prognostic marker include prospective clinical trials and retrospective analysis of archived specimens from a prospective trial.

In evaluating a predictive biomarker, the optimal design is to measure the marker in all patients to identify them as predicted responsive or predicted non-responsive and then to randomize the patients in each group to treatment and control arms.

Simon also discussed a “marker strategy design,” which was referred to as a “prospective screened trial” by Lerman, and agreed that it is often a very inefficient design and that it requires a very large sample size to have sufficient statistical power.

¹ Clinical trial #NCT00310180.

Prospective–Retrospective Study

Some retrospective analyses of archived samples for biomarker studies can result in highly biased conclusions. To address this, Simon and colleagues have proposed a “prospective–retrospective” trial design which uses archived specimens from a single prospective trial to test a specific intended use of an assay and which meets the following criteria:

- (1) adequate amounts of archived tissue must be available from enough patients from an appropriately designed prospective trial (which for predictive factors should generally be a randomized design) for analyses to have adequate statistical power and for the patients included in the evaluation to be clearly representative of the patients in the trial; (2) the test should be analytically and pre-analytically validated for use with archived tissue and the testing should be blinded to the clinical data; (3) the plan for biomarker evaluation should be completely specified in writing before the performance of biomarker assays on archived tissue and should be focused on evaluation of a single completely defined classifier; and (4) the results from archived specimens should be validated using specimens from one or more similar, but separate, studies. (Simon et al., 2009)

Simon also discussed potential revisions to the ASCO LOE scale, which currently classifies retrospective studies as LOE II or lower. He suggested that level 1 evidence could come from either a fully prospective clinical trial or else from two or more prospective–retrospective studies (meeting the proposed criteria above) in which the results were consistent.

In conclusion, analysis of archived tissues for prognostic and predictive biomarkers can provide either a higher or a lower level of evidence in support of clinical utility depending upon several key factors: the analytical validation of the assay; the nature of the study from which the specimens were archived; the number and condition of the specimens; and whether a focused, written plan for analysis of the specified biomarker was developed before assaying any tissue. Studies using archived tissues from prospective clinical trials, when conducted under ideal conditions and independently confirmed, can provide the highest level of evidence (LOE I).

COVERAGE WITH EVIDENCE DEVELOPMENT

The Ontario Model

In 2003 the Ontario Ministry of Health and Long-Term Care implemented a new structure with the goal of implementing an evidence-based approach to policy decision-making regarding medical products and procedures (Figure 3-4) (Goeree and Levin, 2006; Levin et al., 2007). The key component of the new structure is the Ontario Health Technology Advisory

Committee (OHTAC), which receives requests for evidence-based analyses from the Ontario Health System and the Ministry of Health. As Leslie Levin of the Medical Advisory Secretariat (MAS) explained, these requests are passed on to the MAS which coordinates systematic reviews and economic analyses with academic partners. Expert panels are engaged to evaluate the evidence; feedback and input from stakeholders, professionals, the public, and industry is sought; and all evidentiary information is then passed to OHTAC, which develops appraisals based on the evidence and provides recommendations to the Ontario Health System and to the Ministry of Health (Figure 3-4). Adoption of the recommendations can be tracked through a geographic information system.

As one part of the larger structure, a field evaluation program was developed to collect primary data in order to address uncertainties identified in the systematic reviews and to perform post-market assessment of real

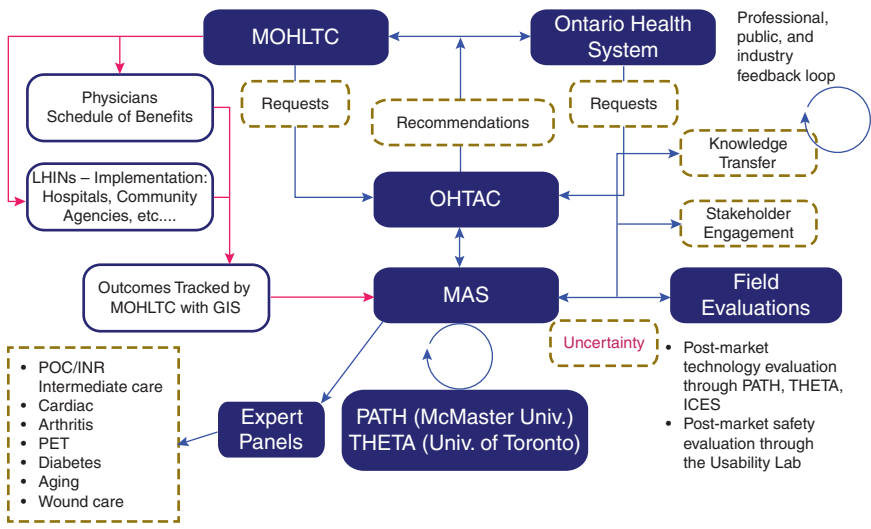


FIGURE 3-4 MAS- and OHTAC-associated structures and linkages. Abbreviations: GIS (geographic information system); ICES (Institute for Clinical Evaluative Sciences); LHIN (local health integration networks); MOHLTC (Ministry of Health and Long-Term Care); PATH (Programs for Assessment of Technology in Health); PET (positron emission tomography); THETA (Toronto Health Economics and Technology Assessment Collaborative). SOURCE: As modified from Levin et al., 2007 by Levin in IOM workshop presentation on November 17, 2010.

world performance of products. The generation and collection of additional data regarding the utilization and impact of a medical intervention as a requirement of a preliminary coverage decision, or coverage with evidence development, comes under the purview of the Ontario Field Evaluation program. Issues that could trigger a field evaluation include, for example, low quality of evidence, incremental net benefit, generalizability questions, and safety issues. Levin noted that 38 field evaluations have been initiated since 2003, with 19 completed thus far. Of the completed studies, 88 percent affected decision making. Ten of these were coverage with evidence development studies. Most of the studies are published in peer-reviewed journals. Levin said that while saving money was not the original intent of the studies, the resulting coverage decisions have contributed to more than \$500 million in cost avoidance.

Examples of Coverage with Evidence Development Field Studies and Recommendations

Levin highlighted several of the ten coverage with evidence development recommendations that have been made thus far (Table 3-2). In response to a published report identifying an increase in restenosis rates for low-risk patients who were treated with bare-metal stents versus those treated with drug-eluting stents, a field evaluation was performed to test the generalizability of this finding. The field study determined that while drug-eluting stents are advantageous for patients at high risk for restenosis, in Ontario there was no significant difference noted for low-risk patients (Tu et al., 2007). As a result, drug-eluting stents are used in only about 30 percent of patients in Ontario, as compared with 90 to 95 percent in the United States. In another case, a systematic review of endovascular abdominal aortic aneurysm repair raised a concern about endoleak (persistent blood flow into the aneurysm sac after the graft procedure). A subsequent prospective field study of 160 patients found that these were primarily Type II endoleaks, which are less serious (Tarride et al., 2008), and an economic analysis showed the procedure is cost effective only for high-risk patients (MAS, 2010). As a result, the decision was made to fund cardiovascular abdominal aortic aneurysm repair only for high-risk patients. MAS has also looked at the clinical utility of positron emission tomography (PET) scanning (Evans et al., 2009). It has insured PET scanning for staging lung cancer, but field studies have not shown clinical utility for head and neck cancer or in staging breast cancer, and these applications of PET are not insured.

As an example of an ongoing coverage with evidence development study, Levin said that MAS is looking at gene expression profiling with Oncotype DX for guiding adjuvant chemotherapy in early breast cancer. As

TABLE 3-2 Summary of Ontario Field Evaluations

Technology (n)	Field Evaluation Overseen by	Type of Study	Reason for Field Evaluation	Result	Policy Decision
Drug eluting stents (DES) (21,000)	PATH, with ICES	Prospective pragmatic registry	Generalizability of RCT evidence and cost effective analysis	Only effective in patients at high risk for restenosis	Funded; 30% conversion from bare-metal to DES (90% in U.S.A.)
Endovascular abdominal aortic aneurysm repair (160)	PATH and single AHSC	Prospective observation	Safety assessment of endoleak	No endoleak; CE only for high surgical risk	Funded for high but not low surgical risk
Multifaceted primary care diabetes program	PATH, with Oxford University	Before-after study using micro simulation economic model	Prioritize investments according to downstream effects and CE following systemic review of diabetes strategy	Most CE were bariatric surgery, MDT; Least, insulin infusion pumps for type II	Bariatric program funded and additional funding for MDT; Insulin infusion pumps for type 2 on hold
64-slice CT angiography (CTA) v coronary angiography (CA) (175)	PATH, with cardiologists, radiologists, selected AHSCs	Patients for CA also underwent CTA	Uncertainty regarding indications for use, CE and QA parameters	Sensitivity lower than reported, reducing CE	OHTAC recommended slow diffusion until sensitivity issue resolved
PET to stage locally advanced NSCLC (310)	OCOG	RCT	Clinical utility in decisions regarding combined modality therapeutics	Terminated by efficacy & safety committee	PET insured for this indication

PET to stage early NSCLC (322)	OCOG	RCT	Resolve inconsistencies to inform decision regarding access	PET reduces futile thoracotomy rates	PET insured for this indication
PET to stage breast cancer (320)	OCOG	Prospective cohort	Compare PET to sentinel lymph node biopsy	No utility in staging	Not insured
PET for colorectal cancer metastatic to liver (400)	OCOG	RCT	Clinical utility in decision for metastatectomy	Accrual completed February 2010	Awaiting results
PET for head and neck cancer (400)	OCOG	Prospective cohort	Clinical utility pre-surgery following radiation therapy	No clinical utility	Not insured
Extracorporeal photopheresis (120)	PATH with AHSC	Prospective observational	Basis for decision regarding funding for GvH and Sezary	Effective in GvH; Inconclusive for Sezary	Insured for GvH; Inconclusive for Sezary - small vol. after backlog dealt with

Abbreviations: AHSC (academic health science center); CE (cost effectiveness); DES (drug eluting stent); GvH (graft vs host); ICES (Institute for Clinical Evaluative Sciences); MDT (multi-disciplinary teams); NSCLC (non-small cell lung cancer); OCOG (Ontario Clinical Oncology Group); OHTAC (Ontario Health Technology Advisory Committee); PATH (Programs for Assessment of Technology in Health); PET (positron emission tomography); QA (quality assurance); RCT (randomized controlled trial).
 SOURCE: As updated from Levin et al., 2007, by Levin in IOM workshop presentation on November 17, 2010.

opposed to the TEC and AHRQ reviews, the MAS evidentiary review found low-quality evidence for its prognostic value and very low-quality evidence for its predictive value in terms of the benefits of a classic chemotherapy regimen. As such, a coverage with evidence proposal has been put forth that would consider three key questions: (1) How does Oncotype DX change treatment? (A prospective cohort study has been proposed.) (2) How does Oncotype DX compare to traditional factors? (Electronically collected data on age, tumor size, grade, ER, PR, and HER-2/neu will make it possible to measure correlations between the Oncotype DX recurrence score and traditional risk classification.) (3) What is the impact of Oncotype DX on breast cancer distant recurrence? (Longitudinal data will be collected.) These studies will be informed by ongoing clinical studies, such as the TAILORx trial being conducted by NCI (Zujewski and Kamin, 2008).

Another example is EGFR mutation testing in non-small cell lung cancer. MAS looked at the predictive value of mutated EGFR based on a retrospective subgroup analysis of archived specimens from a RCT of first-line treatment with gefitinib versus chemotherapy. The results of the analysis suggested a statistically significant improvement in progression-free survival for gefitinib versus chemotherapy in EGFR-mutation-positive patients, but not in EGFR-mutation-negative patients (Zhu et al., 2008). However, results of a similar analysis of second- and third-line chemotherapy (erlotinib versus placebo) were not significant (Shao, 2010), and the studies reviewed were not designed to examine the predictive effects of the mutation. Levin noted that the current pattern of practice is to use erlotinib regardless of EGFR status for second or third-line treatment. As a result, it was recommended that there should be payment for EGFR testing for gefitinib as a first-line treatment and for EGFR testing for erlotinib for second- or third-line treatment, that treatment should be allowed for EGFR-negative patients, but that the response to erlotinib should be monitored by EGFR mutation status and the payment for EGFR testing in this group of patients should be modified based on the findings.

Lessons Learned

In summary, Levin said, coverage with evidence development works, but more efficient methodologies are needed to expedite conclusions. To this end, Levin suggested that evidence-based analysis should be implemented further upstream in the lifecycle of drugs and technologies. Industry, academia, and health systems should be engaged in the premarket phase so that the important evidentiary questions are addressed ahead of time. In this way, it may be possible to influence the development pipeline toward technologies that are more relevant to health systems and to patient outcomes.

CONSTRUCTING CHAINS OF EVIDENCE

The rules of evidence that apply to genomic information are no different than the rules of evidence for other forms of information about prognosis, prediction, or diagnosis, said David Ransohoff of the University of North Carolina at Chapel Hill. A chain of evidence is a series of questions or evidence that together describe the impact of some activity—in this case, a genomic test. A primary issue is what questions should be in the chain.

Analytic Framework

Ransohoff said that established analytic frameworks should be used to develop chains of evidence for genomic tests, and he referred to a presentation that Steven Woolf had given to the roundtable at its March 2010 workshop. Woolf had discussed standard analytic principles that are applied to the evaluation of screening tests regardless of the type of test and had noted that

groups such as the U.S. Preventive Services Task Force and the World Health Organization generally consider five issues when assessing preventative interventions: (1) the burden of suffering from the target condition; (2) the accuracy and reliability of the test; (3) the effectiveness of early detection of the condition; (4) potential harms; and (5) the balance of benefits and harms. (IOM, 2010)

These questions are simple to ask but difficult to answer, Ransohoff said. As discussed by Piper, Calonge, and others, a RCT addressing questions 2 through 5 would be the ideal source of evidence. If there is no clinical trial that can answer all of these questions at once, then evidence must be pieced together.

Evidence about genomic tests is often limited to the accuracy and reliability of the test. However, it is not sufficient for a test to discriminate. The bottom line is the outcome—the benefits and harms that occur because of an intervention choice based on the discriminatory capability of the test. For efficient discovery and development, researchers must work backward from a specific clinical decision and consider benefits, harms, and the intended use and desired features of the test. “Working backwards from a specific clinical scenario is absolutely critical [but] commonly not done,” Ransohoff said.

Clinical Trials

If no RCT has been performed, the evidence is necessarily limited. It is possible to assess the ability of the test to discriminate between popu-

lations but not to determine whether this discrimination or subsequent action affects outcomes. Much of the current genomic evidence is limited to questions of discrimination. Ransohoff also noted that in reviewing the available evidence he found that many of the studies published in clinical journals do not disclose critical details of the study design and participants and sometimes the discrimination observed is actually due to bias or to error, not to biology.

As an example, Ransohoff cited a proteomics study about differential exoprotease activities which was looking to determine whether peptide patterns are sensitive and specific for prostate cancer (Villanueva et al., 2006). In the study the test arm was 100 percent male prostate cancer patients averaging 67 years of age. However, the control group was 58 percent women with a mean age of 35 years, leading to a potential source of bias in the findings. The publication reported this important detail, though only in supplemental data, but many published “-omics” studies are opaque, making it difficult or impossible to assess the strength of the evidence.

Barriers to Implementing an Analytic Framework

An analytic framework model makes clinical sense as an evidentiary pathway, and there is extensive experience with analytical pathways in other fields. The challenge is gathering the evidence to fill out the framework, Ransohoff said. Investigators may not think of data as a product of a study. If the study design is weak, then the link in the evidence chain is also weak. Studies need to be carefully and prospectively designed. Specimens should also be considered a product of a study, and the source of the specimens should be described in detail in the methods section of the publication.

In many cases the rate-limiting step is funding, infrastructure, or informatics, but in generating evidence for genomic test development, the rate-limiting step is formulating the key clinical questions and designing a study that provides strong evidence or a link in the chain. The question that needs to be addressed, Ransohoff said, is whether existing data can be used in a strong design.

Ransohoff also noted that there are opportunities to add well-designed studies onto current practices. As examples, he cited two studies, one a study of prognosis and the other of diagnosis. The prognostic study assessed the five-year risk of developing colon cancer after a negative colonoscopy (Imperiale et al., 2008), while the diagnostic study assessed the ability of colorectal screening to detect advanced proximal neoplasms in asymptomatic adults (Imperiale et al., 2000). Both studies were superimposed on a program that a pharmaceutical company had already implemented as a

clinical benefit for its employees, Ransohoff said, and the prognostic study was done at no additional cost.

Moving Forward

An analytic framework for assessing the impact of a test on outcomes offers an established method for guiding clinical and policy decisions. Conceptualized this way, genetic and genomic information is not exceptional. An RCT to assess the impact of a test on outcome is ideal, but when it is not possible or available, there are other sources of data and evidence that can be used. Banked specimens from clinical trials can be used in prospectively designed studies to address questions about prognosis and prediction, for example, and there are various ways to use other data sources, such as cohort data from a health maintenance organization (HMO). Ransohoff advised participants not to be overly focused on infrastructure, informatics, and data sharing. Rather, the focus should be on answering specific clinical questions and opportunistically designing strong research studies in different settings.

DISCUSSION

Archived Specimens

A participant noted that “you can’t test specimens if you don’t have the specimens to test.” Some fields, such as breast cancer research, collect tissues prospectively when conducting clinical trials, Dan Hayes said. It was noted, however, that the number of patients consenting to this tissue collection has been declining for unknown reasons. Furthermore, collecting other types of samples, such as germ-line DNA, can be very costly. Hayes noted that clinical trials are increasingly run by pharmaceutical companies, which do not necessarily collect and store specimens, and he suggested that the FDA require sponsors of new drug applications to have created specimen banks from their trials, although he acknowledged that intellectual property and other issues would need to be addressed. Simon suggested that, going forward, it will be important to do prospective clinical trials and to store specimens with a prospective–retrospective analysis in mind.

One issue with retrospective sample analysis is that it is not possible to optimize the way in which specimens are acquired for the various purposes that may arise in the wide range of possible future marker investigations. Analyte degradation during storage is another concern. Ransohoff agreed that decay is an issue and cautioned that it is important not to compare newer specimens with older specimens. Researchers need to be aware of the problems that can be caused by such decay, so that bias is not introduced in

the results. “We can be mindful with whatever specimens we have collected in the past and hope to store future ones better,” Ransohoff said. Hayes noted that NCI is developing a prospective systematic funding mechanism to answer some of the more basic questions regarding handling of samples, such as what are the implications if a sample sits for 3 days instead of 3 hours before processing or what happens if the sample is fixed for too long. One approach, Hayes said, is to develop an assay that works in the kind of tissue that is collected and archived currently. Another approach is to develop an assay that is so fundamentally powerful that it will change the way tissue is collected and archived going forward.

Trial Design

Participants discussed what is “clinically relevant.” There are multiple study designs that are valid, and which of them is clinically relevant depends on the particular research question. In the end, the goal is to cross a threshold of evidence based on a combination of observational, retrospective, prospective, and larger, more clinically oriented approaches.

As the panelists noted, many of the studies that are currently being done are not designed to contribute to the evidentiary base in the way that is needed for genomics. A question was raised about how to better train the next generation of clinical investigators to think about biomarker studies. Should the NIH develop some very specific training programs as we move into the genomics era? Ransohoff noted that the system rewards clinicians for getting grants and publishing papers, as opposed to producing products or expanding general knowledge.

Simon said that, in his experience, industry is extremely interested in new clinical trial designs that use predictive biomarkers or candidate predictive biomarkers in new drug development. Industry managers are concerned, however, about what the FDA will require (e.g., prohibitively large clinical trial sizes) and about the potential for more roadblocks in developing new drugs with companion diagnostics. He added that NIH funding is driving much of the basic research on identifying the key targets that could be candidate predictive biomarkers and drug targets.

It was noted that patients are increasingly demanding access to interventions that they regard as essential to their well being and health. Institutions are adopting technologies prematurely, and there is political pressure to approve or cover the latest technologies. This is a knowledge translation problem that needs to be addressed. It is important to consider evidence generation during the premarket phase, as trials are being designed and conducted, before product diffusion into the marketplace.

4

Overcoming Barriers for Evidence Generation

Key Points Raised by Speakers

- Clinical research needs to balance validity with feasibility and timeliness.
- Establishing partnerships and sharing risk among the public sector, payers, and industry will allow for robust development of diagnostic tests with diverse clinical focuses.
- Increased dialogue among stakeholders at various points in the development process could help provide alignment around needed evidence.
- Placing a higher value on diagnostic test and marker development will create incentives to produce the type of evidence needed for decision making.

BALANCING STAKEHOLDER NEEDS

The purpose of comparative effectiveness research is to provide patients, clinicians, and payers with information that is useful in making treatment and coverage decisions. Many, if not most, comparative effectiveness studies will require a conscious decision to sacrifice internal validity in order to increase generalizability, relevance, feasibility, and timeliness, said session moderator Sean Tunis of the Center for Medical Technology Policy. However, this is not something researchers are generally comfortable doing.

The frameworks that have been discussed (e.g., EGAPP, TEC) are designed to maximize internal validity but are not optimal with regard to feasibility or timeliness, which are important to the diagnostics industry and to patients. There is a need for methodologies to evaluate clinical utility that can achieve an acceptable balance of these elements, Tunis said, noting that the correct balance of validity with feasibility and timeliness is not solely a methodology issue. There is also a social judgment that must be made collectively by all stakeholders regarding the acceptable level of uncertainty.

A participant said that the goal in developing the Oncotype DX test for breast cancer, as well as with tests currently in development for colon cancer and prostate cancer, was to gather evidence that would be persuasive to both clinicians and to payers. Performing RCTs for diagnostics is not a necessity, and the length of time they take to produce outcomes data would render the test obsolete. Payer support is also required to ensure patient access.

The participant identified key questions in balancing stakeholder needs, including What are the risks and ramifications of being wrong? and, How comfortable are we with those risks? The further that studies deviate from the principles of the RCTs, the more that certainty declines. To move rapid evaluation forward, new data must be evaluated systematically as they emerge, and decision makers must be willing to stop coverage when it becomes clear that a product does not work as originally thought. It was noted that this is what already happens in many systems, such as the Ontario experience that was presented.¹

PUBLIC-PRIVATE PARTNERSHIPS

One organizational model that can help address issues of funding, knowledge generation, and social change in the area of data sharing is a public-private, pre-competitive research partnership, said Aled Edwards of the Structural Genomics Consortium. Pre-competitive research is knowledge-generating research where data is openly shared and not encumbered by any restrictions on its use. For genetic tests, current pre-competitive research is focused on generating hypotheses.

The Structural Genomics Consortium, founded in 2004, has 250 scientists working in three laboratories located at the University of Toronto, the University of Oxford, and the Karolinska Institute. Initially focused on studying the three-dimensional protein structure of drug targets, the consortium is now also working on pre-competitive medicinal chemistry. Thirty medicinal chemists from industry partners (including GlaxoSmith-Kline, Novartis, Eli Lilly and Company, and Merck) are generating new

¹ Discussed by Levin in Chapter 3.

molecular entities that they are not patenting but rather placing into the public domain without restriction. The pharmaceutical industry, Edwards explained, has numerous potential drug candidates but can only devote resources to fully pursue the very top candidates. By pooling resources, it is possible to assess many more potential drugs and targets. Edwards said that there is now some industry and government interest in funding clinical proof-of-concept trials of public-domain compounds. Data from such trials would also be placed into the public domain without restriction. He suggested that the organizational structure that will be developed to carry out these trials could be used to superimpose genomics studies on the proof-of-concept trials so that they can be carried out at the same time.

Edwards said that the only way to develop a truly robust pipeline of genomic diagnostic tests flowing into the clinic will be to share the risk between the public sector, payers, and industry. The Structural Genomics Consortium was begun by stakeholders, including the pharmaceutical companies, the Wellcome Trust, and the Canadian government, declaring a certain area of scientific research as pre-competitive. The consortium was then given clear milestones and deliverables to reach. Academics were willing to participate because of the no-patent policy, and industry was willing to participate because it could have full access to the information while only putting in a small percentage of the funding. Edwards suggested that much of the early-stage discovery in genetic tests should be done pre-competitively and that failure should be anticipated. This will allow stratification of tests so that not everyone is focusing on the perceived high-value targets and mostly overlooking other potentially important analytes.

As one participant pointed out, the development of partnership frameworks to enable biomarker discovery and development was the subject of a July 2010, Institute of Medicine roundtable activity (IOM, 2011). At that workshop it was noted that the pharmaceutical industry has long been collecting biological specimens from clinical trials and has allowed a number of entities access to those specimens—and to the associated data—for the purposes of developing novel biomarkers and eventually tying them to drug development programs or developing them as stand-alone diagnostics. Several public-private partnerships were given as examples, including the Genomic Applications in Practice and Prevention Network (GAPPNet), Sage Bionetworks, and the biomarkers consortium that is coordinated by the Foundation for the NIH. Pharmaceutical companies left that July workshop willing to enter into collaborations to share data and biospecimens because they understood that they will recoup greater value by partnering than by retaining the information individually. Another point made at the July workshop was that biospecimens collected by the various partners do not necessarily need to be submitted to a central location. They can remain

locally housed at their source as long as they are indexed in some central way so they can be located.

THE EVIDENTIARY BAR FOR CLINICAL UTILITY

Different stakeholders have somewhat different definitions for “evidence of clinical utility.” Robert Epstein of the Medco Research Institute suggested that dialogue between payers and regulators is needed to establish some consistency. Is a difference in health outcomes necessary, and how is that defined? Can a surrogate outcome be assessed, or is a hard outcome necessary? What type of study design is necessary to answer questions of clinical utility? Epstein noted that pharmaceutical product sponsors have an end-of-Phase II meeting with FDA to ensure that they are collecting the necessary data in the appropriate way as they move to the next phase. However, there are no similar meetings with the payers and other stakeholders who ultimately review and help promulgate the use of new technologies. Epstein suggested that it could be very useful to hold a similar mid-development meeting with other stakeholders before the product reaches the market.

It is also important to evaluate the criteria for what constitutes adequate evidence, Epstein said. There are other ways to get data besides RCTs. The computing power and biostatistical expertise available today can reveal many details about a population that were not possible in the past. Epstein urged that “we should begin to look at our criteria and ask ourselves [if] we can improve on them.”

A participant noted that there is much focus on the clinical utility of genomic tests because people believe treatment decisions are based on the results of the tests. But if this is the case, the participant continued, then should not clinical utility be established for all diagnostic tests, given that they all affect patient decision making? Eric Larson of the Group Health Research Institute added that it is often assumed that every new test adds value because it provides information. Genetic testing is not different from other diagnostic testing, and the current focus on genetic tests may provide an opportunity to reframe the overall diagnostics process and highlight the need for the same rigor in developing and evaluating diagnostic tests that are used in developing and evaluating therapeutic products.

DATA SOURCES

Clinical Trials

Conducting a clinical trial for a marketed product, as is done in coverage with evidence development, can be challenging. Citing prostate-specific

antigen (PSA) screening for prostate cancer as an example, Dan Hayes said that so many men were already being screened that it was hard to identify participants for a randomized trial. Another difficulty is that many genetic disorders are relatively rare and, once a marker has been discovered, there are simply not enough patients needed to conduct sufficiently large trials.

Many diagnostic companies are not likely to survive in the current industry environment if they are mandated to conduct prospective trials, said Hayes. The pharmaceutical industry can better afford clinical trials, as the payoff for a new drug can be quite significant compared to research and development costs. Hayes suggested that if markers were more highly valued for their roles in preventing unnecessary treatment of patients who are unlikely to benefit and in identifying those patients who will benefit, it would create an incentive and a revenue stream for diagnostic device developers to carry out prospective trials. Many participants concurred that there is no real incentive to conduct trials on genetic and genomic tests that are not directly tied to a treatment. A participant suggested that perhaps an “Orphan Diagnostic Act” is needed to provide incentives for such tests in the same way that the Orphan Drug Act has provided incentives to develop drugs for rare diseases.

RCTs are artificial relative to real world use of products. Trials have exclusion and inclusion criteria, for example, and they limit or control the concomitant use of other drugs. In practice, anyone can use the product, including off-label use for non-tested indications. One approach that was suggested would be to combine phase III and phase IV studies, extending the traditional phase III trial to assess longer term outcomes in the post-marketing (phase IV) stage.

A participant clarified that the RCT has not really been the standard for diagnostics, which more typically come into practice through technology assessment on chains of evidence. Except for screening tests, Larson added, there are very few published RCTs of diagnostics. It is important to conduct such trials, which are very different from the randomized trials carried out for drugs, in order to make sure that the tests under consideration are actually adding value.

A participant noted that, in reality, genomic medicine is not individually personalized; rather, patients are treated based on the cohort they best fit into. This should guide how evidence is generated. Evaluating the patients and the characteristics of the test and applying it to the proper populations will drive evidence and use more than an RCT.

Coverage with Evidence Development

The health system underinvests in diagnostics, and the coverage with evidence development approach is one way of subsidizing the development

of potentially high-value diagnostics, Tunis said. There are promising technologies that might merit subsidy while they are being further evaluated in order to demonstrate clinical utility. A participant added that there are, however, inherent conflicts between, on the one hand, the purposes of an insurance system and its obligations to its beneficiaries and, on the other hand, what it takes to conduct certain types of trials. There are also issues with covering a product provisionally and then withdrawing coverage. Providers may have made very significant capital investments in equipment and are not likely to give up on these investments easily. In addition, anything that adds to administrative costs is a problem and this will become even more true as health reforms and more stringent standards are implemented. Another issue is that differences in copayments can affect the randomization and blinding processes. Although observational data can be collected feasibly, there are many challenges to superimposing a clinical research structure onto an insurance structure. People are often not willing to volunteer for random selection. One approach that was suggested during the workshop would be to have payers contribute to a pool that supports trials. Employers who purchase coverage for their employees may also be willing to collaborate in and support prospective research, Epstein said.

It was suggested by one participant that the successes and limitations of the coverage with evidence development approach up to now should be evaluated. The longest history of coverage with evidence development in the United States has been with the Medicare system, and it is important to know how successful this has been.

Provisional Approval

While insurers may approve provisional coverage pending the collection of further evidence, a participant from the FDA explained that the agency does not have the legal authority to provide provisional marketing approval for products. Devices must be demonstrated to be safe and effective or else to be substantially equivalent to (i.e., as safe and effective as) an already marketed device. There is no way that provisional device approval could be done in the United States without changes to the existing device law.

In Canada, Levin explained, most of the tests that are provided as an insured service need to have been previously approved, by either Health Canada or by the provincial jurisdiction, in order to be regarded as medically necessary; the Centers for Medicare and Medicaid Services play a similar role in the United States. In Canada, if a genomic test is to be used for targeted therapy, that fact is stated in the licensing approval provided by Health Canada. A workshop participant added that because the health system and insurance companies in Canada are linked, they can fund and

run very large, system-wide trials to develop evidence of utility. This is not possible in the current U.S. health-care system.

Biobanks and Retrospective Studies

While there was much discussion about RCTs, Epstein reiterated that there are other data sources worth considering which may provide a faster way to gather evidence. He suggested, for example, leveraging the data held by closed health-care systems, which have biobanks containing hundreds of thousands of DNA samples from their members. These samples are matched with the patients' electronic health records, which are in turn matched to claims data.

Gregory Germino of the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) agreed with others that a prospective RCT looking at different biomarkers is the gold standard for evidence, but the challenge, he said, is knowing which markers to assess at any given time, and the choice of markers may change over time. To help address this, NIDDK mandates that samples from all large-scale NIDDK-funded studies be turned over to the institute at the end of the study. These samples then become publicly available. Since 2003 NIDDK has collected approximately 73,000 independent DNA samples (with or without cell lines, depending on the nature of the samples) and over 4 million different biospecimens, with the intent of facilitating prospective–retrospective controlled trials. There are no intellectual property issues, and innovators are free to market products that they discover and develop using these samples. Germino added that having a centralized, quality-controlled, and quality-assured repository helps ensure the stability of the samples. NIDDK has found that the research community, after some initial trepidation, has really embraced this repository system.

There are some challenges to conducting studies using biobank samples, Germino noted. Studies may not have been large enough, or may not have enough samples of any given subset, to assess a marker of interest with enough statistical power to draw a strong conclusion based on the study set. A second issue is cost. Two to five percent of the NIDDK clinical trial budget is devoted to the repositories, Germino said, and, while this may seem small, it is actually a substantial draw on the budget. This may also be a factor, he suggested, in why many other institutes have not incorporated biobanks into their study designs.

A question was asked about how applicable the NIDDK biobank approach is to other NIH institutes. Germino responded that the National Heart, Lung, and Blood Institute (NHLBI) has begun a data and biospecimen repository, but there is currently no mechanism to implement a trans-NIH repository.

Not knowing where the biobanks are is certainly a barrier to progress, Germino said. Even within NIH, institutes may not know what sample repositories other institutes have. He suggested that ClinicalTrials.gov should require trial sponsors to identify biobank samples that are linked to their clinical study when they register a trial.

A participant noted that the NCI does not own the specimens from its cooperative trials and has faced some challenges in gaining access to some of the samples. Each cooperative trial group stewards its own specimen bank with NCI oversight, Hayes responded. All specimens from one trial go to a single bank. Requests for sample access are evaluated by peer-review committees composed of study investigators, and there are standardized policies and procedures on how to collect and store specimens. A participant noted that, historically, NCI has provided little support for the collection, maintenance, and distribution of specimens. This expense is often overlooked. Once a request for samples from an outside group is approved, there is no support provided for retrieving all of the specimens, packing them up, and assembling the annotation, data, and statistics. The potential treasure in biobanking clinical trial samples will not be fully realized unless there is sufficient support for acquiring it.

Larson supported merging biobanks with electronic medical records and cited as an example the Electronic Medical Records and Genomics (eMERGE) network organized by the National Human Genome Research Institute (NHGRI), which was designed to facilitate genome-wide association studies (GWASs) in participants from whom phenotypic and environmental data are available through electronic medical records. The HMO Research Network also has a number of sites that have biobanks, such as the Marshfield Clinic in Wisconsin, which has a large population-based biobank. Larson mentioned an inventory of the network biobanks that was expected to be available in December 2010.

A number of participants pointed out that the samples in repositories are precious resources. The model is collaborative while the studies are ongoing and more custodial after they are finished, with precautions in place to allow for the generation of useful information. Germino suggested that one approach to stewardship of specimens is to issue a program announcement and have researchers submit applications to request the samples. A study section would review the applications to ensure that the questions being asked can be addressed with the study design proposed and with the number of samples being requested.

Consent

When done prospectively, it is possible to secure consent for potential future studies that have not yet been envisioned. There are, however,

numerous archival specimens for which consent was not obtained, and a question was raised regarding retrospectively obtaining consent.

Germino said that some sample sets that NIDDK now has from outside institutions were not originally collected with the intention to later transfer the samples to NIDDK for safekeeping and distribution. In those cases, the investigators met with their institutional review boards to discuss whether the original consents would allow subsequent use by NIDDK. These specimens have had identifiers removed so that they are anonymous, and while they are linked to the clinical outcomes data, there are no Health Insurance Portability and Accountability Act identifiers. In every case so far there have been no barriers to having the samples transferred to NIDDK, but Germino noted that this is not the universal experience. Larson added that work from the eMERGE project has found that people who consent to research are likely to re-consent when asked. Forming a partnership with the population that is contributing samples is essential, he said, and “goes a long way to solving these issues around depositing data in large public data sets.”

The Learning Health-Care System

Participants agreed that there is a disconnect between the health system enterprise and the research and industry enterprise. Larson endorsed the Institute of Medicine concept of the “learning health-care system,” which incorporates the generation and application of evidence into the patient care system itself.² Conducting research within existing clinical systems, he said, would presumably be cheaper because the data are already being collected. Such data collected within the context of care are also more likely to be generalizable and applicable to care.

Observational Studies

Epstein drew attention to the Bradford Hill Criteria used in epidemiology to help establish causal relationships, and he suggested that this approach might be helpful in making the field comfortable with the weight of genomic evidence from multiple observational studies in the absence of a RCT.

² See <http://www.iom.edu/Activities/Quality/LearningHealthCare.aspx> for further information about an ongoing IOM consensus study on the learning healthcare system. See also *The Learning Healthcare System: Workshop Summary* available at http://www.nap.edu/catalog.php?record_id=11903.

5

Considerations Moving Forward

Key Points Raised by Speakers

- Developing new models for coverage of tests post-FDA clearance could aid the development of clinical utility evidence.
- Patient management and outcomes should be the focus of intervention development.
- There needs to be greater accounting of and transparency in experimentation for biomarkers.
- The public needs to be educated on the need for evidence of utility for new interventions.

In light of the barriers and complexities associated with the various models of evidence generation discussed and of the need for more timely evidence gathering in order to better meet the needs of patients and providers, the workshop participants considered what the next steps should be in generating new evidence for the development of genomic diagnostic tests.

MEDICAL PRACTICE VERSUS EVIDENCE-BASED RECOMMENDATIONS

Roger Klein, a member of EGAPP, asked why medical practice often diverges from the recommendations issued by evidence-based review groups. For example, there are a number of tests for thrombophilia that

have been cleared by the FDA. These are very commonly ordered assays, yet there is limited evidence that the results of these tests offer any clinically useful information for most people.

Dan Hayes responded that tests are often ordered because they are available and easy to order and there is little cost to the doctor for using them inappropriately. Physicians order marker tests as part of general data gathering about the patient. “A bad marker is as bad as a bad drug,” he said, adding that providers are making critical treatment decisions based on information that may be wrong. Practitioners could do much better in not overtreating patients who will not benefit and in identifying patients who will benefit. This is where marker development is as important as pharmaceutical development, Hayes said. Becker added that the idea of being able to manage clinical effect (e.g., the hoped-for effect against a tumor or the range of adverse events) is something that lies in the realm of clinical practice. It is a matter of judgment and experience. The question in the mind of a clinician deciding whether to adopt a new test is, What does it add to what is already done in day-to-day practice? Hayes added that clinicians are sometimes of the opinion that practice guidelines are made without considering individual patients and the nuances that come with each. Such contextual issues are difficult to build into the guidelines.

One approach that could affect practice would be for third-party payers who agreed to cover a genomic diagnostic test to then deny coverage of a treatment if the provider treated the patient differently than what the test result had directed (but only if the recommendations from the results are clear cut). Blue Cross and Blue Shield plans, for example, do not pay for trastuzumab treatment for a HER2-negative patient because such a test result indicates it is not a medically appropriate treatment. However, some participants said, this approach does raise concerns about patient autonomy and about patients’ rights to change their minds on treatment course in certain circumstances.

HETEROGENEITY OF EVIDENCE-BASED DECISIONS

Moderator Sharon Terry of Genetic Alliance pointed out the variation in evidence-based decisions. For example, FDA recently cleared the breast cancer recurrence test, MammaPrint, for more broad use in all age groups, while the Blue Cross and Blue Shield Association TEC determined that Oncotype DX met its criteria but MammaPrint did not, and EGAPP has concluded there is insufficient evidence to recommend for or against either test.

Piper responded that the Oncotype DX test was reviewed by TEC several times, and the final vote to cover the test was very close. Similarly,

the final EGAPP recommendation was not arrived at easily. (As discussed by Calonge in Chapter 3, the insufficient evidence conclusion for Oncotype DX was classified as “encouraging indirect evidence.”) These nuances do not come across because only the final decision or recommendation is released. TEC assessments summarize the available evidence, Piper noted, but are not a practice guideline, while EGAPP is able to incorporate some of these contextual issues. This difference, she said, may add to some of the disparities between recommendations.

Hayes said that the ASCO committee was widely criticized and accused of being United States–centric for having recommended the 21-gene test, Oncotype DX, which was developed by a company in the United States, and not recommending the 70-gene test, MammaPrint, from an Amsterdam-based company. Hayes explained that ASCO’s Oncotype DX decision was based on the review of studies that asked clinically relevant questions regarding node-negative, estrogen receptor–positive women and tamoxifen that were applicable to practice. Becker said that one of the most important differences between the tests was the fact that Oncotype DX claimed to be able to therapeutically manage patients, while MammaPrint could not specifically make that claim. It came down to clinical utility versus clinical validity.

A test that is cleared or approved by FDA may not necessarily be reimbursed by payers, Becker noted. Clinical validation ties the test to the disease and diagnosis, which may lead to a prognosis, but the payer is focused on the utility. This system, where regulatory approval is disassociated from reimbursement approval, has been in place for a long time in the United States, and, Becker said, “It is not a system which was tailored for the rapidly evolving circumstances that we see now.” A participant warned against linking FDA clearance and reimbursement too closely, as this may lead to other problems. “Either the bar at the FDA would be too high for the commerce aspect or too low for what a rational system should be spending its resources on.”

Currently, a device developer might allot three to five years in its expected development timeline to navigate the FDA clearance process, a participant said. If the target is changed to clinical utility, that could easily push this phase of development out to 7 to 10 years. If payers will reimburse for devices after they are cleared or approved by the FDA, it gives the manufacturer an opportunity both to sell the product and to prepare and conduct clinical utility trials. Developing biospecimen banks will help close the gap between what FDA requires and what third-party payers require. Coverage with evidence development is a valuable tool as well, as it allows for provisional coverage while studies are being conducted.

RESEARCH

Piper reiterated that research is too often driven by the desire to secure grants, publications, FDA approval, and reimbursement, when the real focus should be whether the intervention will make a difference in managing the patient. This should be the starting point in choosing tests to evaluate.

It was suggested by several participants that the Institute of Medicine roundtable could help define the top 10 genomics clinical questions that need to be answered. Such a list would not have to be the definitive list, but it could provide well-justified examples and help flesh out other questions that would also be appropriate. It will also be important to compare these questions with the priorities of the medical system at large, one participant commented. The point was made again that the success of Oncotype DX was due to the fact that the clinical question came first, then the developer figured out what kind of dataset was needed to answer the question and found the most appropriate previously conducted clinical trial. As another participant noted, there needs to be a focus on addressing the salient clinical questions.

With regard to building the genomics evidence infrastructure, one participant suggested starting with such a list of ten different, clinically important questions and assembling groups to decide how best to answer those questions and to determine the infrastructure that would need to be developed to facilitate that approach. Some of the answers will come from analysis of biobanked samples, but there may also be ways to answer questions outside of clinical trials. It would be very informative to solve a set of specific, important problems over the next five to ten years and see what evidentiary approaches emerge.

A participant pointed out that research has changed over time as NIH has required grantees to include minorities, children, and women in studies and has required that data-sharing plans be devised. What if NIH required grantees to demonstrate the potential for translation into practice?

Another participant observed that device manufacturers tend to design trials that place the product into a lower category for FDA clearance so that the product can be brought to market more quickly. While it may speed up approval, this approach does not deliver information about the true value of the test for patients and clinicians.

It was noted that one of the themes that has emerged in every roundtable workshop is the idea of having a framework that allows for access to high-quality biological samples connected to clinical data and which facilitates discovery and validation of genetic and genomic biomarkers. “We have to build the sandbox that everybody will play in,” said a participant. It was suggested that GAPPNet (a stakeholders’ group formed by CDC)

could offer a forum for taking discussions and ideas from the roundtable workshops and fleshing out how to translate them into practice by pulling together public and private partners to conduct pilot projects to test the ideas in different contexts.

BROAD CONSIDERATIONS

One participant offered several observations relevant to moving forward with an evidence generation model for genomic tests:

- Individual learning curves vary considerably, and learning is an ongoing process that extends beyond medical/health professional school.
- Part of the process is learning to be comfortable with being wrong.
- Variability among both consumers and providers of health care results in different messages being given at different times and in different situations.
- Familiarity with new tests is variable (e.g., knowing what actions to take in the face of strong test results).
- There is a general lack of comfort with and understanding of change.
- There is a lack of comfort with ambiguity.
- There is no perfect evidence and no absolute truth. We need to set a standard of excellence that providers can be comfortable with.

TRANSPARENCY

A participant said that the evidence base is damaged by selective publication of clinical trials. Although manufacturers are required to register all clinical trial protocols publicly in ClinicalTrials.gov, the participant said she believed that not all trial results are being reported.

Hayes agreed that ClinicalTrials.gov has been a great step forward. The Consolidated Standards of Reporting Trials have helped to ensure transparency in clinical trial reporting, as have the requirements by many journals that authors provide details about basic research studies, such as the specific reagents used. But for some reason, the translational area in between basic and clinical research has been left behind. Hayes suggested that many journal editors, who are generally very careful about not accepting clinical trials that are not registered in ClinicalTrials.gov or that do not provide essential experimental details, have abrogated their responsibility in terms of biomarkers, because markers are not valued in the way that clinical trials are. Hayes mentioned a new effort to initiate a registry for tumor marker

studies. This registry would help ensure that studies are being conducted with prospectively written protocols and that negative results come to light.

It was further suggested that the roundtable engage major journal editors and urge them to pay more attention to tumor marker studies with regard to design, quality, and transparency.

PUBLIC EXPECTATIONS

Participants discussed the need for a public education effort regarding genomic interventions. Terry said that many people believe the FDA to be primarily interested in the public's health and not in commerce and assume that agency clearance implies that the product is beneficial and should be adopted. On the other hand, a participant noted, the public often perceives FDA as being too slow or hindering access to products (such as happened with products targeting the human immunodeficiency virus). Another participant added that the current environment is one of "immediate-communications" and that health literacy and public understanding about evidence are highly variable. While eager for new products to be available on the market, the public in general does not demand evidence to support claims about new interventions. As a participant said, "If we could simply educate the public to ask [for evidence] every time they hear a claim, I think it would help us." It was also noted that whole-genome analysis will enable providers to give patients their entire genome, but it will not be of much practical use unless the genome can be related to various kinds of clinical evidence.

6

Final Remarks

The primary goals of this workshop were to discuss the types of evidence needed by the various stakeholders involved with genomic diagnostics and to identify mechanisms to obtain high-quality evidence more efficiently. From the regulatory perspective, decisions by the FDA to clear or approve medical devices (including genomic diagnostic tests) for marketing are based on the safety and effectiveness of the product. From the payer perspective, demonstration of clinical utility or evidence of improved health outcomes is required for decisions to provide coverage. Evidence-based review groups also take contextual issues into account and look for overall net benefit (the balance of benefits or potential benefits versus harms or potential harms) when making a recommendation for or against the use of a genomic application. From a provider perspective, the focus is on value for the patient and on improving individual outcomes by identifying the most appropriate treatment for a person's situation (e.g., increasing the chances of cure, survival, or palliation or decreasing exposure to toxicity from unnecessary or inappropriate therapy).

With these perspectives and those given by other stakeholders during the workshop in mind, participants reviewed a variety of approaches to evidence generation, such as clinical trials, retrospective analysis of archived specimens, coverage with evidence development, and chains of evidence and analytic frameworks, and offered ideas and strategies that could help move evidence generation for genomic diagnostic test development forward.

CHAIR'S SUMMARY

Workshop chair Debra Leonard concluded the workshop by highlighting key questions and topics for further discussion (perhaps in future workshops facilitated by the IOM roundtable) and potential action items identified from the discussions.

Evidence

- Consider how to close the gap between FDA and payers' evidence requirements. (The FDA, payers, evidence-based review groups, and providers all expressed a willingness to come together with IOM facilitation for further discussion.)
- Conduct an analysis of the cost-effectiveness of an “analytic framework” process (constructing a chain of evidence from pieces that together provide adequate supporting data) versus conducting one or two high-quality, prospective randomized controlled clinical trials.
- Define what constitutes “adequate evidence.” Perfect evidence is unattainable. What level of certainty will allow the transition of a genomic intervention into clinical practice?
- Educate the public on the need for evidence to support clinical tests and clinical practice. Create a public demand for evidence and reduce the demand for tests simply because they are available or new.

Reimbursement and Coverage

- Discuss new economic/reimbursement models that place value on tests that can help identify when a particular treatment will not be beneficial and thereby prevent unnecessary costly therapeutic interventions.
- Discuss implementation of a system that does not pay for treatment if a prognostic or predictive genomic test is available and the results of that test do not support treatment in the patient.

Medical Practice

- Consider whether safety and efficacy (as determined by FDA review and approval/clearance) is sufficient to support the clinical use of a new genomic test in the context of medical practice relative to an individual patient's situation or whether large amounts of population data should be required.

- Explore the disconnect between medical practice and evidence-based review recommendations: Why do physicians tend to ignore evidence-based review recommendations and how can uptake of recommendations be enhanced?

Clinically Focused Research

- Foster a patient-centric research system to focus diagnostic test research on clinically important questions.
- Develop a cooperative arena for identifying the top 10 clinically important questions and the resources or mechanisms to generate that evidence collaboratively. Consider convening a large annual meeting to analyze the available data that can be used to answer clinically significant questions.
- Discuss with journal editors the importance of transparency in the reporting of diagnostic test validation studies toward establishing a strong, unbiased evidence base. Journals should publish only validation studies that meet study design quality criteria, and results should be published regardless of whether the outcome is positive or negative.

Access to Clinical Trial Specimens and Data

- Establish a single index of annotated clinical trial specimens and closed health systems (e.g., Medco) that have the ability to conduct genomic test development projects. Explore whether the mechanism used by GAPPNet could be used to achieve this goal.
- Consider adopting across all institutes of the NIH the central repository model of NIDDK, which requires that specimens from NIDDK-funded clinical studies be submitted to NIDDK's sample and data repositories, thereby facilitating controlled storage conditions and resource sharing.
- Continue to develop ClinicalTrials.gov to be more complete with regard to trial information posted on that site and, in particular, to facilitate the reporting of trials with negative outcomes.
- Develop models of data sharing for genomic tests and test development.

Academia

- Engage academic medical center leadership in discussions about how faculty contributions are valued; change the academic promotion and reward systems to more highly value and to better

reward clinically important research outcomes and collaborative efforts, rather than rewarding solely on the number of grants and publications.

- Develop a link between academic research and development and the health-care teams at academic medical centers, with the shared goal of improving the health of patients. Identify the incentives for aligning these stakeholders in order to leverage specific clinical experiences and develop novel research initiatives.
- Train clinical investigators in diagnostic test development and in study design options and optimization.

References

- Baselga, J., D. Tripathy, J. Mendelsohn, S. Baughman, C. C. Benz, L. Dantis, N. T. Sklarin, A. D. Seidman, C. A. Hudis, J. Moore, P. P. Rosen, T. Twaddell, I. C. Henderson, and L. Norton. 1999. Phase II study of weekly intravenous trastuzumab (herceptin) in patients with HER2/neu-overexpressing metastatic breast cancer. *Seminars in Oncology* 26(4 Suppl 12):78-83.
- Benowitz, N. L., O. F. Pomerleau, C. S. Pomerleau, and P. Jacob, 3rd. 2003. Nicotine metabolite ratio as a predictor of cigarette consumption. *Nicotine and Tobacco Research* 5(5):621-624.
- Berry, D. A., K. A. Cronin, S. K. Plevritis, D. G. Fryback, L. Clarke, M. Zelen, J. S. Mandelblatt, A. Y. Yakovlev, J. D. Habbema, and E. J. Feuer. 2005. Effect of screening and adjuvant therapy on mortality from breast cancer. *New England Journal of Medicine* 353(17):1784-1792.
- Berry, D. A., C. Cirincione, I. C. Henderson, M. L. Citron, D. R. Budman, L. J. Goldstein, S. Martino, E. A. Perez, H. B. Muss, L. Norton, C. Hudis, and E. P. Winer. 2006. Estrogen-receptor status and outcomes of modern chemotherapy for patients with node-positive breast cancer. *Journal of the American Medical Association* 295(14):1658-1667.
- Brass, E. P. 2010. The gap between clinical trials and clinical practice: The use of pragmatic clinical trials to inform regulatory decision making. *Clinical Pharmacology & Therapeutics* 87(3):351-355.
- de Jong, A. E., Y. M. Hendriks, J. H. Kleibeuker, S. Y. de Boer, A. Cats, G. Griffioen, F. M. Nagengast, F. G. Nelis, M. A. Rookus, and H. F. Vasen. 2006. Decrease in mortality in Lynch syndrome families because of surveillance. *Gastroenterology* 130(3):665-671.
- Evans, W. K., A. Laupacis, K. Y. Gulenchyn, L. Levin, and M. Levine. 2009. Evidence-based approach to the introduction of positron emission tomography in Ontario, Canada. *Journal of Clinical Oncology* 27(33):5607-5613.
- Goeree, R., and L. Levin. 2006. Building bridges between academic research and policy formulation: The PRUFE framework—an integral part of Ontario's evidence-based HTPA process. *Pharmacoeconomics* 24(11):1143-1156.

- Gonzales, D., S. I. Rennard, M. Nides, C. Oncken, S. Azoulay, C. B. Billing, E. J. Watsky, J. Gong, K. E. Williams, and K. R. Reeves. 2006. Varenicline, an $\alpha 4\beta 2$ nicotinic acetylcholine receptor partial agonist, vs sustained-release bupropion and placebo for smoking cessation: A randomized controlled trial. *Journal of the American Medical Association* 296(1):47-55.
- Harris, L., H. Fritsche, R. Mennel, L. Norton, P. Ravdin, S. Taube, M. R. Somerfield, D. F. Hayes, and R. C. Bast, Jr. 2007. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *Journal of Clinical Oncology* 25(33):5287-5312.
- Imperiale, T. F., D. R. Wagner, C. Y. Lin, G. N. Larkin, J. D. Rogge, and D. F. Ransohoff. 2000. Risk of advanced proximal neoplasms in asymptomatic adults according to the distal colorectal findings. *New England Journal of Medicine* 343(3):169-174.
- Imperiale, T. F., E. A. Glowinski, C. Lin-Cooper, G. N. Larkin, J. D. Rogge, and D. F. Ransohoff. 2008. Five-year risk of colorectal neoplasia after negative screening colonoscopy. *New England Journal of Medicine* 359(12):1218-1224.
- IOM (Institute of Medicine). 2010. *The value of genetic and genomic technologies: Workshop summary*. Washington, DC: The National Academies Press.
- IOM. 2011. *Establishing precompetitive collaborations to stimulate genomics-driven product development: Workshop summary*. Washington, DC: The National Academies Press.
- Jarvinen, H. J., M. Aarnio, H. Mustonen, K. Aktan-Collan, L. A. Aaltonen, P. Peltomaki, A. D. L. Chapelle, and J. P. Mecklin. 2000. Controlled 15-year trial on screening for colorectal cancer in families with hereditary nonpolyposis colorectal cancer. *Gastroenterology* 118(5):829-834.
- Lee, J. J., and E. Chu. 2007. First-line use of anti-epidermal growth factor receptor monoclonal antibodies in metastatic colorectal cancer. *Clinical Colorectal Cancer* 6 (Suppl 2):S42-S46.
- Levin, L., R. Goeree, N. Sikich, B. Jorgensen, M. C. Brouwers, T. Easty, and C. Zahn. 2007. Establishing a comprehensive continuum from an evidentiary base to policy development for health technologies: The Ontario experience. *International Journal of Technology Assessment in Health Care* 23(3):299-309.
- Malaiyandi, V., S. D. Goodz, E. M. Sellers, and R. F. Tyndale. 2006. CYP2A6 genotype, phenotype, and the use of nicotine metabolites as biomarkers during ad libitum smoking. *Cancer Epidemiology, Biomarkers & Prevention* 15(10):1812-1819.
- Marchionni, L., R. F. Wilson, S. S. Marinopoulos, A. C. Wolff, G. Parmigiani, E. B. Bass, and S. N. Goodman. 2008. Impact of gene expression profiling tests on breast cancer outcomes. Evidence report/technology assessment no. 160. (prepared by the Johns Hopkins University Evidence-Based Practice Center under contract no. 290-02-0018). *Agency for Healthcare Research and Quality Publication No. 08-E002* January.
- MAS (Medical Advisory Secretariat). 2010. *Endovascular repair of abdominal aortic aneurysms in low surgical risk patients: Rapid review*. Toronto, ON: Medical Advisory Secretariat. Available at http://www.health.gov.on.ca/english/providers/program/mas/tech/rapid/pdf/rr_evar_20100113.pdf (accessed March 24, 2011).
- Moss, A. J., W. Shimizu, A. A. Wilde, J. A. Towbin, W. Zareba, J. L. Robinson, M. Qi, G. M. Vincent, M. J. Ackerman, E. S. Kaufman, N. Hofman, R. Seth, S. Kamakura, Y. Miyamoto, I. Goldenberg, M. L. Andrews, and S. McNitt. 2007. Clinical aspects of type-1 long-QT syndrome by location, coding type, and biophysical function of mutations involving the KCNQ1 gene. *Circulation* 115(19):2481-2489.
- Paik, S., S. Shak, G. Tang, C. Kim, J. Baker, M. Cronin, F. L. Baehner, M. G. Walker, D. Watson, T. Park, W. Hiller, E. R. Fisher, D. L. Wickerham, J. Bryant, and N. Wolmark. 2004a. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *New England Journal of Medicine* 351(27):2817-2826.

- Paik, S., S. Shak, G. Tang, C. Kim, J. Baker, M. Cronin, R. Baehner, M. Walker, D. Watson, J. Bryant, and N. Wolmark. 2004b. Risk classification of breast cancer patients by the recurrence score assay: Comparison to guidelines based on patient age, tumor size, and tumor grade. Paper read at 27th Annual San Antonio Breast Cancer Symposium, December 8-11, 2004, San Antonio, TX.
- Paik, S., G. Tang, S. Shak, C. Kim, J. Baker, W. Kim, M. Cronin, F. L. Baehner, D. Watson, J. Bryant, J. P. Costantino, C. E. Geyer, Jr., D. L. Wickerham, and N. Wolmark. 2006. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *Journal of Clinical Oncology* 24(23):3726-3734.
- Palomaki, G. E., M. R. McClain, S. Melillo, H. L. Hampel, and S. N. Thibodeau. 2009. EGAPP supplementary evidence review: DNA testing strategies aimed at reducing morbidity and mortality from Lynch syndrome. *Genetics in Medicine* 11(1):42-65.
- Pencina, M. J., R. B. D'Agostino, Sr., R. B. D'Agostino, Jr., and R. S. Vasan. 2008. Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. *Statistics in Medicine* 27(2):157-172; discussion 207-212.
- Pepe, M. S., H. Janes, G. Longton, W. Leisenring, and P. Newcomb. 2004. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *American Journal of Epidemiology* 59(9):882-890.
- Peto, R., J. Boreham, M. Clarke, C. Davies, and V. Beral. 2000. UK and USA breast cancer deaths down 25% in year 2000 at ages 20-69 years. *Lancet* 355(9217):1822.
- Ray, R., R. F. Tyndale, and C. Lerman. 2009. Nicotine dependence pharmacogenetics: Role of genetic variation in nicotine-metabolizing enzymes. *Journal of Neurogenetics* 23(3):252-261.
- Roden, D. M. 2008. Clinical practice. Long-QT syndrome. *New England Journal of Medicine* 358(2):169-176.
- Schmeler, K. M., H. T. Lynch, L. M. Chen, M. F. Munsell, P. T. Soliman, M. B. Clark, M. S. Daniels, K. G. White, S. G. Boyd-Rogers, P. G. Conrad, K. Y. Yang, M. M. Rubin, C. C. Sun, B. M. Slomovitz, D. M. Gershenson, and K. H. Lu. 2006. Prophylactic surgery to reduce the risk of gynecologic cancers in the Lynch syndrome. *New England Journal of Medicine* 354(3):261-269.
- Shao, Y. Y. 2010. Gefitinib or erlotinib in the treatment of advanced non-small cell lung cancer. *Discovery Medicine* 49(June 2010). Available at <http://www.discoverymedicine.com/Yu-Yun-Shao/2010/06/19/gefitinib-or-erlotinib-in-the-treatment-of-advanced-non-small-cell-lung-cancer/> (accessed March 25).
- Simon, R. M., S. Paik, and D. F. Hayes. 2009. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *Journal of the National Cancer Institute* 101(21):1446-1452.
- Tarride, J. E., G. Blackhouse, G. De Rose, T. Novick, J. M. Bowen, R. Hopkins, D. O'Reilly, and R. Goeree. 2008. Cost-effectiveness analysis of elective endovascular repair compared with open surgical repair of abdominal aortic aneurysms for patients at a high surgical risk: A 1-year patient-level analysis conducted in Ontario, Canada. *Journal of Vascular Surgery* 48(4):779-787.
- TEC (Blue Cross and Blue Shield Association Technology Evaluation Center). 2005. Gene expression profiling for managing breast cancer treatment. *Technology Evaluation Center Assessment Program* 20(3):1-5.
- TEC. 2007. Epidermal growth factor receptor mutations and tyrosine kinase inhibitor therapy in advanced non-small-cell lung cancer. *Technology Evaluation Center Assessment Program* 22(6):1-22.
- TEC. 2008a. Gene expression profiling of breast cancer to select women for adjuvant chemotherapy. *Technology Evaluation Center Assessment Program* 22(13):1-8.

- TEC. 2008b. Genetic testing for long-QT syndrome. *Technology Evaluation Center Assessment Program* 22(9):1-5.
- TEC. 2011. Epidermal growth factor receptor mutations and tyrosine kinase inhibitor therapy in advanced non-small-cell lung cancer. *Technology Evaluation Center Assessment Program* 25(6):1-30.
- Teutsch, S. M., L. A. Bradley, G. E. Palomaki, J. E. Haddow, M. Piper, N. Calonge, W. D. Dotson, M. P. Douglas, and A. O. Berg. 2009. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) initiative: Methods of the EGAPP working group. *Genetics in Medicine* 11(1):3-14.
- Tu, J. V., J. Bowen, M. Chiu, D. T. Ko, P. C. Austin, Y. He, R. Hopkins, J. E. Tarride, G. Blackhouse, C. Lazzam, E. A. Cohen, and R. Goeree. 2007. Effectiveness and safety of drug-eluting stents in Ontario. *New England Journal of Medicine* 357(14):1393-1402.
- Villanueva, J., D. R. Shaffer, J. Philip, C. A. Chaparro, H. Erdjument-Bromage, A. B. Olshen, M. Fleisher, H. Lilja, E. Brogi, J. Boyd, M. Sanchez-Carbayo, E. C. Holland, C. Cordon-Cardo, H. I. Scher, and P. Tempst. 2006. Differential exoprotease activities confer tumor-specific serum peptidome patterns. *Journal of Clinical Investigation* 116(1):271-284.
- Wolff, A. C., M. E. Hammond, J. N. Schwartz, K. L. Hagerty, D. C. Allred, R. J. Cote, M. Dowsett, P. L. Fitzgibbons, W. M. Hanna, A. Langer, L. M. McShane, S. Paik, M. D. Pegram, E. A. Perez, M. F. Press, A. Rhodes, C. Sturgeon, S. E. Taube, R. Tubbs, G. H. Vance, M. van de Vijver, T. M. Wheeler, and D. F. Hayes. 2007. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *Journal of Clinical Oncology* 25(1):118-145.
- Zhu, C. Q., G. da Cunha Santos, K. Ding, A. Sakurada, J. C. Cutz, N. Liu, T. Zhang, P. Marrano, M. Whitehead, J. A. Squire, S. Kamel-Reid, L. Seymour, F. A. Shepherd, and M. S. Tsao. 2008. Role of KRAS and EGFR as biomarkers of response to erlotinib in National Cancer Institute of Canada Clinical Trials Group Study BR.21. *Journal of Clinical Oncology* 26(26):4268-4275.
- Zujewski, J. A., and L. Kamin. 2008. Trial assessing individualized options for treatment for breast cancer: The TAILORx trial. *Future Oncology* 4(5):603-610.

A

Workshop Agenda

Generating Evidence for Genomic Diagnostic Test Development:
A Workshop

November 17, 2010

The Keck Center, Room 100
500 Fifth Street, N.W.
Washington, DC 20001

MEETING OBJECTIVE

To explore the strengths and limitations of the evidence that is being generated for the clinical validity and utility of genomic diagnostic tests.

- What evidence is required from stakeholders?
- How is evidence currently being generated?
- Are there innovative ways to generate higher-quality evidence more efficiently?
- What are the barriers to generating this evidence and how can they be overcome?

7:45 – 8:30 A.M. **WORKING BREAKFAST**

8:30 – 8:45 A.M. **PUBLIC WORKSHOP BEGINS**

8:30 – 8:45 A.M. **WELCOME AND INTRODUCTORY REMARKS**
*Debra Leonard, Professor and Vice Chair for
Laboratory Medicine, Director of the Clinical
Laboratories, Director of the Pathology
Residency Training Program, Weill Cornell
Medical Center of Cornell University*

8:45 – 10:50 A.M. EVIDENCE

Moderator: *Roger Klein, Director, Molecular Oncology Laboratory at BloodCenter of Wisconsin and Clinical Assistant Professor of Pathology at the Medical College of Wisconsin*

8:45 – 9:05 A.M. FDA Perspective

Robert L. Becker, Jr., Chief Medical Officer for the Office of In Vitro Diagnostic Device Evaluation and Safety, Center for Devices and Radiological Health, U.S. Food and Drug Administration

9:05 – 9:25 A.M. Payer Perspective

Margaret Piper, Director of Genomics Resources for the Blue Cross and Blue Shield Technology Evaluation Center

9:25 – 9:45 A.M. Evidence-Based Review Group Perspective

Ned Calonge, President and CEO of the Colorado Trust

9:45 – 10:05 A.M. Health-Care Provider Perspective

Dan Hayes, Clinical Director of the Breast Oncology Program and Stuart B. Padnos Professor in Breast Cancer Research, University of Michigan Comprehensive Cancer Center

10:05 – 10:35 A.M. Panel Discussion**10:35 – 10:50 A.M. BREAK****10:50 A.M. – 12:55 P.M. NEW MODELS FOR EVIDENCE GENERATION: APPLIED SCIENCE**

Moderator: *Elizabeth Mansfield, Director of the Personalized Medicine Staff, Office of In Vitro Diagnostic Device Evaluation and Safety, Center for Devices and Radiological Health, U.S. Food and Drug Administration*

- 10:50 – 11:10 A.M. Randomized Clinical Trials and Practical Clinical Trials in Pharmacogenomics**
Caryn Lerman, Mary W. Calkins Professor and Director of the Tobacco Use Research Center; Deputy Director of the Abramson Cancer Center, University of Pennsylvania
- 11:10 – 11:30 A.M. Utilizing Archived Samples to Generate New Evidence**
Richard Simon, Chief of the Biometric Research Branch, National Cancer Institute
- 11:30 – 11:50 A.M. Coverage with Evidence Development**
Leslie Levin, Head of the Medical Advisory Secretariat and Senior Medical, Scientific and Health Technology Advisor for the Ministry of Health and Long-Term Care, Ontario, CANADA
- 11:50 A.M. – 12:10 P.M. Constructing Chains of Evidence**
David F. Ransohoff, Professor, Departments of Medicine and Epidemiology, University of North Carolina at Chapel Hill
- 12:10 – 12:55 P.M. Discussion**
- 12:55 – 1:55 P.M. WORKING LUNCH**
- 1:55 – 4:20 p.M. OVERCOMING BARRIERS FOR EVIDENCE GENERATION**
 Moderator: *Sean Tunis, Director of the Center for Medical Technology Policy*
- 1:55 – 2:35 P.M. Panelist Remarks**
Aled Edwards, Director and CEO of the Structural Genomics Consortium; Professor, Department of Medical Biophysics, University of Toronto
Robert S. Epstein, Chief Medical Officer and President of the Medco Research Institute
Gregory G. Germino, Deputy Director of the National Institute for Diabetes, Digestive, and Kidney Diseases

Eric Larson, Executive Director of the Group Health Research Institute

2:35 – 3:25 P.M. **Panel Discussion**

3:25 – 4:05 P.M. **Roundtable and Audience Discussion**

4:05 – 4:20 P.M. **BREAK**

4:20 – 5:20 P.M. **STRATEGIES FOR MOVING FORWARD**
 Moderator: *Sharon Terry, President and Chief Executive Officer of Genetic Alliance*

4:20 – 5:20 P.M. **Stakeholder Reaction Panelists**

Robert L. Becker, Jr., Chief Medical Officer for the Office of In Vitro Diagnostic Device Evaluation and Safety, Center for Devices and Radiological Health, U.S. Food and Drug Administration

Dan Hayes, Clinical Director of the Breast Oncology Program and Stuart B. Padnos, Professor in Breast Cancer Research, University of Michigan Comprehensive Cancer Center

Roger Klein, Director of the Molecular Oncology Laboratory at BloodCenter of Wisconsin and Clinical Assistant Professor of Pathology at the Medical College of Wisconsin

Margaret Piper, Director of Genomics Resources for the Blue Cross and Blue Shield Technology Evaluation Center

5:20 – 5:50 P.M. **SUMMARY AND CONCLUSIONS**

5:20 – 5:50 P.M. **Review and Conclusions**
Debra Leonard, Professor and Vice Chair for Laboratory Medicine, Director of the Clinical Laboratories, Director of the Pathology Residency Training Program, Weill Cornell Medical Center of Cornell University

Appendix B

Speaker Biographical Sketches

Robert L. Becker, Jr., M.D., Ph.D., is chief medical officer for the Office of In Vitro Diagnostic Devices Evaluation and Safety (OIVD), Center for Devices and Radiological Health (CDRH), at the U.S. Food and Drug Administration, with special attention to inter-office coordination on regulation of newly emerging genetic and genomic IVDs. Dr. Becker previously served as director of the Division of Hematology and Immunology Devices in OIVD. He is experienced in regulation of IVDs aimed at cell- and tissue-based specimens (e.g., classical hematology, flow cytometry, cytology, and histopathology) as well as of blood coagulation tests and immunoserologic tests. Dr. Becker earned his M.D. and Ph.D. in immunology at Duke University, and he is board certified in anatomic and clinical pathology. He served in the U.S. Air Force as a pathologist at the Armed Forces Institute of Pathology, Washington, DC, from 1983 to 2004, specializing in urologic pathology and with research and clinical service applying image analysis and flow cytometry to diagnostic pathology.

Ned Calonge, M.D., M.P.H., is the president and chief executive officer of the Colorado Trust, a philanthropic foundation dedicated to advancing the health and well-being of the people of Colorado. He is an associate professor of family medicine at the Colorado School of Medicine, University of Colorado, Denver, and an associate professor of epidemiology at the Colorado School of Public Health. Outside of the trust, Dr. Calonge is a member and past president of the Colorado Medical Board, which licenses and regulates physicians. He teaches epidemiology, biostatistics and research methods at the University of Colorado Schools of Medicine and Public

Health. He is a member of the Delta Dental Foundation Board and chairs the board of directors for LiveWell, Colorado. Nationally, Dr. Calonge is the chair of the United States Preventive Services Task Force and a member of the Centers for Disease Control and Prevention's (CDC's) Task Force on Community Preventive Services. He is the chair of the CDC's Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Workgroup and is a member of the Secretary's Advisory Committee on Heritable Disorders in Newborns and Children in the Maternal and Child Health Bureau in the Health Resources and Services Administration. Dr. Calonge received his B.A. in chemistry from the Colorado College, his M.D. from the University of Colorado and his M.P.H. from the University of Washington; he is board certified in both family medicine and preventive medicine. Prior to coming to the trust, Dr. Calonge was the chief medical officer of the Colorado Department of Public Health and Environment.

Aled Edwards, Ph.D., is Banbury Professor of Medical Research at the University of Toronto, visiting professor of chemical biology at the University of Oxford, and chief executive of the Structural Genomics Consortium (SGC), an Anglo-Canadian-Swedish public-private partnership created to increase substantially the number of protein structures of relevance to human health available in the public domain, without restriction on use. Funded by industry, governments, and charitable foundations, the SGC accounts for more than a quarter of the world's output of human protein structures and more than 75 percent of the world's output of proteins from the parasites that cause malaria, toxoplasmosis, and cryptosporidiosis. Dr. Edwards believes that the discovery of new medicines would be most efficiently accomplished by performing many aspects of drug discovery research, from discovery to clinical proof of concept, within pre-competitive research consortia, and by de-emphasizing the perceived value of patents. Dr. Edwards was scientific consultant for the Canadian dramatic TV series, *ReGenesis*, and has founded a number biotechnology companies.

Robert S. Epstein, M.D., M.S., is chief medical officer and president, Medco Research Institute, Medco Health Solutions, Inc. Dr. Epstein joined Medco in 1995 and has served as chief medical officer since 1997. In this capacity, he is responsible for formulary development, clinical guidelines, drug information services, accreditation oversight, and personalized medicine services. He is also responsible for analysis and reporting for Medco's clients. In 2009, Dr. Epstein was named president of the Medco Research Institute, where he oversees Medco's peer-reviewed research initiatives and collaborations in the areas of personalized medicine, comparative effectiveness, and chronic conditions. Dr. Epstein was trained as an epidemiologist and worked in public health and academia

before joining the private sector. He is a past elected president of the International Society of Pharmacoeconomics and Outcomes Research and has served on the board of directors for the Drug Information Association. In 2008, Dr. Epstein was nominated and elected to the stakeholder group of the Evaluation of Genomic Applications in Practice and Prevention of the Centers for Disease Control and Prevention as well as the Agency for Healthcare Research and Quality Centers for Education and Research on Therapeutics steering committee. He has published more than 50 peer-reviewed medical articles and book chapters and serves as a reviewer for several influential medical journals.

Gregory G. Germino, M.D., is the deputy director of the National Institute of Diabetes and Digestive and Kidney Diseases, a senior investigator in its intramural program, and an adjunct professor of medicine at the Johns Hopkins School of Medicine. He received his medical degree from the Pritzker School of Medicine at the University of Chicago in 1983 and pursued clinical training in internal medicine and nephrology at Yale–New Haven Hospital. He spent a research year in the Nuffield Department of Medicine at Oxford University in the department of Sir David Weatherall before returning to Yale to complete his research training in nephrology. He joined the faculty of the Johns Hopkins University School of Medicine in 1992 and became a professor in the Department of Medicine and in the Department of Molecular Biology and Genetics in 2003. He was an affiliate member of the McKusick-Nathans Institute of Genetic Medicine from 2002 to 2009. Dr. Germino moved to the National Institutes of Health in 2009 to assume his current position. Dr. Germino served on the scientific advisory board of the Polycystic Kidney Research Foundation from 1994 to 2000 and was a councilor of the American Society of Clinical Investigation from 2004 to 2006 as well as a member of the board of directors of the Federation of American Societies for Experimental Biology from 2004 to 2009. His research interests are in genetic renal disease.

Daniel Fleming Hayes, M.D., is the clinical director of the breast oncology program at the University of Michigan Comprehensive Cancer Center (UM CCC), where he is the Stuart B. Padnos Professor of Breast Cancer Research. He received a bachelor's degree (1974) and a master's degree (1977) from Indiana University. He received his M.D. from the Indiana University School of Medicine in 1979, followed by a residency in internal medicine at the University of Texas Health Science Center at Dallas (Parkland Memorial and affiliated hospitals). He served a fellowship in medical oncology from 1982 to 1985 at Harvard's Dana Farber Cancer Institute (DFCI) in Boston. In 1992, he assumed the role of the medical director of the Breast Evaluation Center at DFCI. He held that title until

1996, when he moved to the Georgetown University Lombardi Cancer Center. In 2001, Dr. Hayes joined the UM CCC and continues treating patients and doing research in translational science. Dr. Hayes and colleagues published the first reports concerning the development of the CA15-3 blood test, which is currently used worldwide to evaluate patients with breast cancer. He has become an internationally recognized leader in the use of this and other tumor markers, such as HER-2, circulating tumor cells, and pharmacogenomics. In 2007, he was awarded the American Society of Clinical Oncology's Gianni Bonadonna Breast Cancer Award. He is chair of the Breast Cancer Translational Medicine Committee of the Southwest Oncology Group and chair of the Correlative Sciences Committee of the U.S. Breast Cancer Intergroup, and he co-chairs the Expert Panel for Tumor Marker Practice Guidelines for the American Society of Clinical Oncology.

Roger D. Klein, M.D., J.D., is director of the Molecular Oncology Laboratory at the BloodCenter of Wisconsin's Diagnostic Laboratories. He has focused on DNA- and RNA-based testing for evaluation of cancer patients. He has helped to expand the BloodCenter's services to diagnose and treat patients with blood-related cancers. Previously, Dr. Klein worked at the H. Lee Moffitt Cancer Center, where he served as medical director of molecular diagnostics, and was an assistant professor in the Department of Oncologic Sciences at the University of South Florida Medical School. Dr. Klein's academic and clinical efforts focus on the translation of molecular genetics knowledge into clinical diagnostic tests. In addition, Dr. Klein has an active research program involving the ethical, legal, and social implications of the Human Genome Project, with particular emphasis on the areas of intellectual property and the regulation of *in vitro* diagnostics and clinical laboratories. Dr. Klein is a member of the Evaluation of Genomic Applications in Practice and Prevention working group of the Centers for Disease Control and Prevention and a consultant to the Clinical and Molecular Genetics Advisory Panel of the U.S. Food and Drug Administration. He also serves on the College of American Pathologists' Molecular Oncology Resource Committee, and the Association for Molecular Pathology's professional relations and CPT coding committees. Dr. Klein has a B.A. degree in chemistry from Case Western Reserve University, magna cum laude, where he was a member of Phi Beta Kappa and Phi Alpha Theta. He has an M.D. from Case Western Reserve University and earned a J.D. degree from Yale Law School, where he was an Olin Fellow and served as an articles editor of the *Yale Journal on Regulation*.

Eric Larson, M.D., M.P.H., MACP, is executive director of the Group Health Research Institute. A graduate of Harvard Medical School, he trained in internal medicine at Beth Israel Hospital in Boston, completed

a Robert Wood Johnson Clinical Scholars and masters of public health program at the University of Washington, and then served as chief resident of University Hospital in Seattle. He served as medical director of the University of Washington Medical Center and associate dean for clinical affairs from 1989 to 2002. His research spans a range of general medicine topics and has focused on aging and dementia, including a long running study of aging and cognitive change set in the Group Health Cooperative—The UW/Group Health Alzheimer’s Disease Patient Registry/Adult Changes in Thought Study. He has served as president of the Society of General Internal Medicine, chair of the Office of Technology Assessment (OTA)/Department of Health and Human Services Advisory Panel on Alzheimer’s Disease and Related Disorders, and chair of the board of regents of the American College of Physicians (2004–2005). He is an elected member of the Institute of Medicine of the National Academy of Sciences.

Debra Leonard, M.D., Ph.D., received her M.D. and Ph.D. from the New York University School of Medicine, and is currently professor and vice chair for laboratory medicine in the Department of Pathology and Laboratory Medicine, and director of the clinical laboratories for New York–Presbyterian Hospital’s Cornell campus (NYPH–WCMC). She is also director of the pathology residency training program at NYPH–WCMC. Dr. Leonard was previously director of molecular pathology at the University of Pennsylvania School of Medicine and is a nationally recognized expert in molecular pathology. She has served on several national committees that develop policy for the use of genetic and genomic technologies and information, including most recently the Secretary’s Advisory Committee on Genetics, Health and Society that advises the Secretary of Health and Human Services. Dr. Leonard is editor of two molecular pathology textbooks and has spoken widely on various molecular pathology test services, the future of molecular pathology, and the impact of gene patents on molecular pathology practice. Dr. Leonard is interested in the use of genomic technologies in the practice of medicine to improve patient outcomes.

Caryn Lerman, Ph.D., is the Mary W. Calkins Professor in the Department of Psychiatry and the Annenberg Public Policy Center, and the interim director of the Abramson Cancer Center at the University of Pennsylvania. She also directs Penn’s Center for Interdisciplinary Research on Nicotine Addiction. Dr. Lerman’s work focuses on the translation of research in genetics, pharmacology, and neuroscience to develop and improve treatments for nicotine addiction. She was among the first to publish evidence for genetic influences on smoking behavior and to study genetic modifiers of response to pharmacotherapy for nicotine addiction. The ultimate goal of her research program is to translate these findings to clinical practice

in the form of personalized medicine for smoking cessation. Dr. Lerman is a member of the Institute of Medicine. Additional honors include the American Psychological Association Award for Outstanding Contributions to Health Psychology, the American Society of Preventive Oncology Cullen Award for Tobacco Research, and the Alton Ochsner Award for Research Relating Tobacco and Health. She is currently the president of the Society for Research on Nicotine and Tobacco. Dr. Lerman serves on the National Institute on Drug Abuse Advisory Council, and is a former member of the National Cancer Institute board of scientific advisors, and the National Advisory Council for Human Genome Research.

Leslie Levin, M.B., M.D., FRCP, FRCPC, is the senior medical, scientific and health technology advisor to the Ministry of Health and Long-Term Care (MOHLTC) and head of the Medical Advisory Secretariat (MAS), which is mandated to provide evidentiary platforms for policy decision making. In this capacity, Dr. Levin has overall leadership in evidence-based assessment relating to all health technologies, including equipment, devices, medical and surgical interventions, and health systems. In these initiatives Dr. Levin works closely with the leadership of Academic Health Science Centers, academia, and industry. Dr. Levin was instrumental in creating the Ontario Health Technology Advisory Committee, which advises the MOHLTC on the adoption of all non-drug health technologies. He was a member of the Canadian Task Force on Health Technology. Dr. Levin initiated the Cancer Care Ontario evidence-based cancer guidelines initiative as vice president of Cancer Care Ontario, and he was instrumental in creating a unique evidence-based provincial cancer drug program. He has published in the area of health technology assessment, evidence-based analysis, and chemotherapy dose intensity research in addition to numerous other cancer research publications. He is a member of the Health Technology Assessment Council of the International Society of Pharmacoeconomics and Clinical Outcomes, a member of the medical advisory panel of Blue Cross and Blue Shield (U.S.), and a director on the board of the International Network of Agencies for Health Technology Assessment. Dr. Levin has senior leadership experience in public and hospital health and in academic administration in addition to experience in clinical and wet bench research. Dr. Levin has advised governments and funding agencies on evidence-based policy decision-making in Canada, the United States, Scotland, Australia, and China and has forged academic collaborations internationally. Dr. Levin is a professor in the Department of Medicine, University of Toronto, and is a senior consultant in medical oncology at the Princess Margaret Hospital. Prior to this, he was professor and chair of the Department of Oncology at the University of Western Ontario and chief executive officer of the London Regional Cancer Centre. His M.D. was awarded by the University

of Birmingham for research in cancer immunology, and he has Royal College certification in internal medicine.

Elizabeth Mansfield, Ph.D., is the director of the personalized medicine staff in the Office of In Vitro Diagnostic Device Evaluation and Safety (OIVD) at the Center for Devices and Radiological Health at the U.S. Food and Drug Administration (FDA), where she is developing a program to address companion and novel diagnostic devices. She was previously a senior policy analyst in OIVD, managing policy and scientific issues. Dr. Mansfield served as the director of regulatory affairs at Affymetrix, Inc. from 2004 to 2006. She previously served in other positions at FDA, including scientific reviewer and genetics expert. Dr. Mansfield received her Ph.D. from Johns Hopkins University and completed further postdoctoral training at the National Cancer Institute and the National Institute for Arthritis, Musculoskeletal, and Skin Diseases.

Margaret Piper, Ph.D., M.P.H., is the director of genomics resources at the Blue Cross and Blue Shield Association Technology Evaluation Center (TEC, www.bcbs.com/tec), an evidence-based practice center funded by the Agency for Healthcare Research and Quality (AHRQ–EPC). She has been with TEC since 1994, joining the staff full-time in 1999. Her experience at TEC has focused on systematic reviews of medical technology, including topics in autoimmunity and transplantation, oncology, laboratory medicine, and genomics/genetic testing. Dr. Piper has authored over 30 TEC systematic reviews and reports and has co-authored 4 AHRQ–EPC reports. Among other outreach activities, Dr. Piper has served on the Medicare Evidence Development & Coverage Advisory Committee of the Centers for Medicare and Medicaid Services and on a work group for the Institute for Quality in Laboratory Medicine, and she currently serves on the working group for the Evaluation of Genomic Applications in Practice and Prevention (EGAPP) project of the Centers for Disease Control and Prevention (CDC). Roles of the EGAPP working group include establishing methods and process for evidence-based evaluation of genetic tests, prioritizing and selecting topics for review, participating in technical expert panels for commissioned evidence reviews, and developing conclusions or recommendations based on the evidence. In addition to these activities, Dr. Piper has given presentations on evidence-based evaluation of genetic tests at meetings organized by the Institute of Medicine, AHRQ, and the National Cancer Institute (NCI). Prior experience includes over 13 years of managing a variety of clinical diagnostic laboratory departments in both academic hospital and commercial clinical laboratory settings, designing and evaluating new laboratory diagnostics for the biomedical industry, consulting with physicians, publishing, and volunteer teaching for professional organiza-

tions in laboratory medicine. In 2000, Dr. Piper received a distinguished service award from the American Society of Clinical Pathologists Commission on Continuing Education. Following a mid-career NCI fellowship in cancer prevention and control, which included obtaining an M.P.H. in epidemiology, Dr. Piper gained experience in cancer epidemiology at the NCI and subsequently at the CDC, with a focus on cancer genetics. Dr. Piper has a B.S. in molecular biology (University of Wisconsin–Madison), a Ph.D. in immunology (Duke University), and an M.P.H. in epidemiology (Emory University).

David F. Ransohoff, M.D., is an internist (gastroenterology, University of Chicago) and clinical epidemiologist (research methodology, Yale University) with long-standing interest in improving the methods used to evaluate diagnostic tests, particularly for screening. He and clinical epidemiology mentor Alvan Feinstein wrote one of the first papers about methodologic challenges in evaluating diagnostic tests. After years of research on conventional tests for cancer and other diseases, he extended his work, starting 10 years ago, to include molecular tests, after being introduced to this area by EXACT Sciences (stool DNA markers for colon cancer) and by a sabbatical supported by the National Cancer Institute (NCI). At the University of North Carolina at Chapel Hill, Dr. Ransohoff directs a faculty development program in translational research, and he works closely with staff and investigators in NCI's Division of Cancer Prevention and other NCI groups on topics related to molecular markers for cancer. His areas of expertise include clinical epidemiology, methodology for evaluating diagnostic tests, cancer screening, and internal medicine.

Richard Simon, D.Sc., is chief of the Biometric Research Branch (BRB) of the National Cancer Institute (NCI) (<http://linus.nci.nih.gov/brb>), where he is chief statistician for the Division of Cancer Diagnosis and Treatment. He holds a doctoral degree in applied mathematics and computer science from Washington University in St. Louis, Missouri. With over 450 publications, he has developed statistical methods widely used today in clinical trials. Dr. Simon is an elected member of the American Statistical Association, a former member of the National Research Council Committee on Theoretical and Applied Statistics, and a former member of the Oncologic Drug Advisory Committee of the U.S. Food and Drug Administration. He is a member of the editorial board of several journals in the areas of cancer research and bioinformatics. In 1998, Dr. Simon established the Molecular Statistics and Bioinformatics Section of the NCI, a multi-disciplinary group of scientists developing and applying methods for the application of genomics data to cancer therapeutics. He is involved in the training of statistical and computational scientists in trans-disciplinary research. He is the archi-

tect of BRB-ArrayTools software for the analysis of microarray expression and copy number data; with over 10,000 registered users in 65 countries, it has been cited in over 1,000 publications.

Sharon Terry, M.A., is president and chief executive officer of the Genetic Alliance, a coalition of over 600 disease specific advocacy organizations working to increase capacity in advocacy organizations and to leverage the voices of the millions of individuals and families affected by genetic conditions. She is the founding executive director of PXE International, a research advocacy organization for the genetic condition pseudoxanthoma elasticum (PXE). She is at the forefront of consumer participation in genetics research, services, and policy and serves as a member of many of the major governmental advisory committees on medical research, including the Cellular, Tissue and Gene Therapies Advisory Committee of the U.S. Food and Drug Administration and the Secretary's Advisory Committee on Heritable Disorders in Newborns and Children. She is a member of the board of directors of the Biotechnology Institute and is on the advisory board of the Johns Hopkins Genetics and Public Policy Center funded by the Pew Charitable Trusts. She is the chair of the Coalition for Genetic Fairness, composed of advocates, health-care providers, and industry, which is working to enact effective federal policy to prohibit genetic information discrimination. She is also chair of the Social Issues Committee of the American Society of Human Genetics. In 2005, she received an honorary doctorate from Iona College for her work in community engagement and haplotype mapping. Ms. Terry is a co-founder of the Genetic Alliance Biobank and serves as president of its board. It is a centralized biological and data [consent/clinical/environmental] repository catalyzing translational genomic research on rare genetic diseases. The biobank works in partnership with academic and industrial collaborators to develop novel diagnostics and therapeutics to better understand and treat these diseases. Along with the other co-inventors of the gene associated with PXE (ABCC6), she holds the patent for the invention. She codirects a 19-lab research consortium and manages 52 offices worldwide for PXE International.

Sean Tunis, M.D., M.Sc., is the founder and director of the Center for Medical Technology Policy in Baltimore, Maryland. Dr. Tunis was a member of the Institute of Medicine Committee on Initial National Priorities for Comparative Effectiveness Research. He advises a wide range of domestic and international public and private healthcare organizations on issues of comparative effectiveness, evidence-based medicine, clinical research, reimbursement, and health technology policy. Through September 2005, Dr. Tunis was the director of the Office of Clinical Standards and Quality

and chief medical officer at the Centers for Medicare and Medicaid Services (CMS). In this role he had lead responsibility for clinical policy and quality for the Medicare and Medicaid programs, which provide health coverage to over 100 million U.S. citizens. Dr. Tunis supervised the development of national coverage policies, quality standards for Medicare and Medicaid providers, quality measurement and public reporting initiatives, and the Quality Improvement Organization program. As chief medical officer, Dr. Tunis served as the senior advisor to the CMS administrator on clinical and scientific policy. He also co-chaired the CMS Council on Technology and Innovation. Dr. Tunis joined CMS in 2000 as the director of the coverage and analysis group. Before joining CMS, Dr. Tunis was a senior research scientist with the Technology Assessment Group, where his focus was on the design and implementation of prospective comparative effectiveness trials and clinical registries. Dr. Tunis also served as the director of the health program at the congressional Office of Technology Assessment and as a health policy advisor to the U.S. Senate Committee on Labor and Human Resources, where he participated in policy development regarding pharmaceutical and device regulation. He received a B.S. degree in biology and history of science from the Cornell University School of Agriculture and a medical degree and masters in health services research from the Stanford University School of Medicine. Dr. Tunis did his residency training at UCLA and the University of Maryland in emergency medicine and internal medicine. He is board certified in internal medicine and holds adjunct faculty appointments at Johns Hopkins University, Stanford University, and the University of California San Francisco schools of medicine.

C

Registered Attendees

Kevin Anderson
DHS Science & Technology

George Annas
Boston University

Naomi Aronson
Blue Cross and Blue Shield
Technology Evaluation Center

LeeAnn Bailey
Deloitte LLP

Robert Becker
U.S. Food and Drug
Administration

Judith Benkendorf
American College of Medical
Genetics

Jim Black
TRICARE Management Activity

Bruce Blumberg
Kaiser Permanente

Denise Bonds
National Heart, Lung, and Blood
Institute

Khaled Bouri
U.S. Food and Drug
Administration

Joel V. Brill
Independent

Amy Brower
American College of Medical
Genetics

Jonca Bull
Novartis Pharmaceuticals

Ned Calonge
The Colorado Trust

Kathryn Camp

Office of Dietary Supplements
National Institutes of Health

Julie Cantor-Weinberg

College of American Pathologists

A. Egon Cholakian

IRDF Project Harvard / Columbia

Melina Cimler

Illumina Inc.

Ralph Coates

Centers for Disease Control and
Prevention

Constanze Coon

Deloitte LLP

Marsha Deckman

Peconic Bay Medical Center

Patricia Deverka

Center for Medical Technology
Policy

Melinda DiVito

The National Academies

Bernard Edelman

Vietnam Veterans of America

Aled Edwards

Structural Genomics Consortium

Robert Epstein

Medco Health Solutions, Inc.

William Gregory Feero

National Human Genome Research
Institute

Amanda Field

Genetic Alliance

Andrew Freedman

National Cancer Institute

Gregory Germino

National Institute of Diabetes and
Digestive and Kidney Diseases

Geoffrey Ginsburg

Duke University

Theresa Gorenc

Avalere Health

Linda Griffith

University of Kansas Medical
Center

Manjit Hanspal

National Institutes of Health

Daniel Hayes

University of Michigan
Comprehensive Cancer Center

Peter Henderson

PWHenderson

Bradford Hirsch

Duke University

India Hook-Barnard

The National Academies

Janet Jenkins-Showalter

Roche

Sharon Kardia

University of Michigan School of
Public Health

Alice Kau

National Institute of Child Health
and Human Development

Alisha Keehn

American College of Medical
Genetics

James Kelly

Roche Molecular Systems

Mohamed Khan

American Medical Association

Muin Khoury

Centers for Disease Control and
Prevention

Roger Klein

BloodCenter of Wisconsin

Pei Koay

Chemical Heritage Foundation

Hanns Kuttner

Hudson Institute

Eric Larson

Group Health Research Institute

Meghan Lawson

Cepheid, Inc.

Teresa Lee

Advanced Medical Technology
Association

Greg Lennon

SNPedia

Debra Leonard

College of American Pathologists

Caryn Lerman

University of Pennsylvania

Leslie Levin

Medical Advisory Secretariat,
Ministry of Health & Long-
Term Care

Hallie Lewis

Cepheid, Inc.

Adriana Malheiro

National Center for Biotechnology
Information

Elizabeth Mansfield

U.S. Food and Drug
Administration

Robert McCormack

Veridex, LLC.

Stephen Michael

Policy Directions Inc.

James Mills

National Institutes of Health

C. Douglas Monroe

Kaiser Permanente

Amy Muhlberg

Life Technologies

Sharon Murphy

Institute of Medicine

Kathryn Phillips

University of California, San
Francisco

Andre Pilon

American Association for Cancer
Research

Margaret Piper

Blue Cross and Blue Shield
Technology Evaluation Center

Irene Prabhu Das

National Cancer Institute

Jennifer Puck

University of California, San
Francisco

Elizabeth Rach

Institute of Medicine

David Ransohoff

University of North Carolina at
Chapel Hill

Lynn Reid

Boston University School of
Medicine

Wendy Rubinstein

NorthShore University Health
System

Debi Sarkar

Maternal and Child Health Bureau
Health Resources and Services
Administration

Thomas Scarnecchia

Digital Aurora

Sheri Schully

National Cancer Institute

Fay Shamanski

College of American Pathologists

Richard Simon

National Cancer Institute

Scott Steele

University of Rochester

Robin Stompler

Auburn Health Strategies, LLC

Fang Sun

ECRI Institute

Weimin Tang

ShanghaiBio (US)

Cara Tenenbaum

Ovarian Cancer National Alliance

Sharon Terry

Genetic Alliance

Barry Thompson

American College of Medical
Genetics

Martha Turner

American Nurses Association

Paul Vetter

PerkinElmer

Michael Watson

American College of Medical
Genetics

Daniel J. Wattendorf

Department of the Air Force

Meredith Weaver
American College of Medical
Genetics

John Yeh
Beth Israel Deaconess Medical
Center/Harvard Medical
School

Catherine Wicklund
National Society of Genetic
Counselors

