

EFFECTIVENESS GUIDANCE DOCUMENT



Evaluation of Clinical Validity and Clinical Utility of Actionable Molecular Diagnostic Tests in Adult Oncology

Release Date: May 1, 2013

BETTER EVIDENCE.
BETTER DECISIONS.
BETTER HEALTH.

Center for Medical Technology Policy
World Trade Center Baltimore
401 E Pratt Street, Suite 631
Baltimore, Maryland 21202

410.547.CMTP (P)
410.547.5088 (F)
info@cmtpNet.org (E)
@cmtp_baltimore (Tw)

www.cmtpNet.org

EFFECTIVENESS GUIDANCE DOCUMENT WRITING TEAM

Patricia Deverka, MD, MS, MBE

Senior Research Director
Center for Medical Technology Policy

Donna Messner, PhD

Research Director
Center for Medical Technology Policy

Tania Dutta, MS, MPP

Project Manager
Center for Medical Technology Policy

ACKNOWLEDGEMENTS

The writing team would like to thank the following for their help in developing the content of this Effectiveness Guidance Document:

Sean Tunis, MD, MSc

President and Chief Executive Officer
Center for Medical Technology Policy

Krissy Krackowsky, MPH, MBA

Program Director
Ulman Cancer Fund for Young Adults

Brad Hirsch, MD, MBA

Assistant Professor
Duke Cancer Institute & Duke Clinical Research Institute

FUNDING

Support for this Effectiveness Guidance Document was provided by the Well Point Foundation, Kaiser Permanente National Community Benefit Fund at the East Bay Community Foundation, United Health Foundation as well as CMTP's unrestricted funds. The following members of the Molecular Diagnostics Advisory Group also generously provided financial support: AstraZeneca, Sanofi-Aventis; Millennium: A Takeda Oncology Company; GlaxoSmithKline; Becton Dickinson, Daiichi Sankyo, Cepheid, Genentech, Amgen, Genomic Health, Myriad Genetics, PricewaterhouseCoopers.

ACKNOWLEDGEMENTS

The development of this Effectiveness Guidance Document (EGD) was supported throughout the process by a Technical Working Group (TWG) including researchers, methodologists, clinicians, product developers, patient advocates, and payers. As described in the preface, the participation of individuals from these organizations does not constitute endorsement.

The Molecular Diagnostics Advisory Group (MDAG) provided technical and methodological input to ensure applicability of the recommendations for diagnostic test development.

The project team would like to thank Lisa McShane of the Biometric Research Branch, Division of Cancer Treatment and Diagnosis, National Cancer Institute, for her statistical and methodological advice and contributions.

We would also like to acknowledge the intellectual contributions from the public and private health insurance industry including Elaine Jeter, Palmetto GBA; Jeffrey Roche, Centers for Medicare & Medicaid Services (CMS); Lee Newcomer, United Healthcare, Alan Rosenberg, WellPoint; Michael Doherty, Kaiser Permanente and Edmund Pezalla, Aetna, who provided critical insights for development of this EGD.

The research team conducted key informant interviews to identify key evidence gaps, as well as major methodological and implementation issues that would affect the clinical integration of molecular diagnostic (MDx) tests in Oncology. The list of interviewees is provided in Appendix A.

The project team would also like to acknowledge the following CMTP staff members for their dedication and hard work to make this project successful: Corinne Warren, Project Manager; Ellen Tambor, Research Manager; Merianne Spencer, Research Associate; Julie Simmons, Manager, Marketing & Communications; Janelle King, Administrative Assistant; and Caitlin Rose, Administrative Assistant. Finally, we would like to thank Scott Allocco, President of SJA Healthcare Strategies, for his support on this project.

MOLECULAR DIAGNOSTICS TECHNICAL WORKING GROUP

**Gary H. Lyman, MD, MPH, FASCO, FACP,
FRCP(Edin)**

Professor of Medicine and Director, Comparative
Effectiveness and Outcomes Research Program
Duke University School of Medicine and the Duke
Cancer Institute

Robert McCormack, PhD

Director of Technology Assessment
Veridex, LLC (a Johnson & Johnson company)

David Nelson, PhD

President and Chief Executive Officer
Epic Sciences, Inc.

Margaret Piper, PhD, MPH

Director, Genomics Resources
BCBSA Technology Evaluation Center

After March 28, 2013

Senior Investigator
Kaiser Permanente, The Center for Health
Research

Richard Simon, DSc

Chief, Biometric Research Branch, Division of
Cancer Treatment and Diagnosis
National Cancer Institute

David Parkinson, MD

Venture Partner
New Enterprise Associates, Inc.

Howard McLeod, PharmD

Director, UNC Institute for Pharmacogenomics and
Individualized Therapy
University of North Carolina, Chapel Hill

Louis Jacques, MD

Director, Coverage & Analysis Group
Centers for Medicare & Medicaid Services

Linda Bradley, PhD

Associate Director, Division of Medical Screening
and Special Testing
Department of Pathology and Laboratory Medicine
Women and Infants Hospital RI
Research Associate Professor
Alpert Medical School of Brown University

Mary Lou Smith, JD, MBA

Co-Founder
Research Advocacy Network

MOLECULAR DIAGNOSTICS ADVISORY GROUP

Robert A. Beckman, MD

Executive Director, Clinical Development Oncology
Daiichi Sankyo Pharmaceutical Development

Donald A. Bergstrom, MD, PhD

Associate VP and Global Head, Translational and
Experimental Medicine
Sanofi, Oncology

Jeff Bush, CT(ASCP), MBA

Director, Worldwide Reimbursement &
Reimbursement Policy
Becton, Dickinson and Company

After February 2013

Vice President, Payer Markets & Reimbursement
AssureRx Health

Lorrie Carr, MBA

Senior Director, US Market Access
Millennium: A Takeda Oncology Company

After November 2012

Divisional VP, Specialty & Infusion Sales
Walgreens

Nicholas Ferenc, MS, MBA

Director
Sanofi Oncology

Sheikh Usman Iqbal, MD, MPH, MBA

Head of Oncology, Global Evidence & Value
Development
Sanofi

Theresa LaVallee, PhD

Senior Director, Translational Medicine
MedImmune

Ash Malik, MS, MBA

Director
PricewaterhouseCoopers (PwC)

Rajiv Mallick, PhD

Director HEOR
Daiichi Sankyo Inc

Gerald McDougall

Partner,
PricewaterhouseCoopers (PwC)

Matthew Monberg, MS

Manager, US Health Outcomes
GlaxoSmithKline

William E. Murray

Vice President, Intellectual Property - Oncology
Cepheid

Scott D. Patterson, PhD

Executive Director, Medical Sciences
Amgen Inc.

Alissa Pereira, MPH

Senior Manager, Public Policy and Reimbursement,
Government Affairs
Genentech Inc.

Steven L. Richardson, MD, MSc

Director, Medical Affairs, Managed Care
Genomic Health Inc.

Greg Rossi, PhD

VP Payer Evidence
Astrazeneca

MOLECULAR DIAGNOSTICS ADVISORY (cont'd)

Rupa Roychowdhury, PhD

US Medical Affairs Pathfinder,
Immunotherapeutics BU
GSK
Until December 2012

Michele Schoonmaker, PhD

Vice President of Government Affairs
Cepheid

Art Small, MD

Director Oncology Outcomes Research US Medical
Affairs
Genentech Inc.

William Trepicchio, PhD

Senior Director, Translational Medicine,
Millennium Pharmaceuticals: A Takeda Oncology
Company

Richard J. Wenstrup, MD

Chief Medical Officer
Myriad Genetic Laboratories, Inc.

ABOUT CMTP

The Center for Medical Technology Policy (CMTP) is an independent, non-profit 501(c)(3) organization that aims to make health care more effective and affordable by improving the quality, relevance, and efficiency of health care research. We focus on the design and implementation of comparative effectiveness research to produce information that helps patients, clinicians, and payers make informed treatment and policy decisions. CMTP provides a trusted forum in which a broad range of stakeholders can collaborate to identify important research questions, design appropriate studies, and develop innovative partnerships to implement these studies.

In an environment of increasing cost pressures, we are experts in bringing together diverse perspectives to create evidence that leads to better health care, while sustaining meaningful innovation. In particular, CMTP applies advanced methods of stakeholder engagement in three main program areas:

Comparative Effectiveness Research Standards

CMTP defines and publishes methodological standards and guidance for CER and PCOR that reflect the information needs of patients, clinicians, and payers. We also advise public and private sector researchers on real-world study designs that reflect these standards.

Research Infrastructure

CMTP develops technically sophisticated and highly implementable processes and products that enhance the clinical research enterprise. This includes facilitating the creation of large data collection systems, research networks and registries and streamlining informed consent methods in community-based studies.

Policy

CMTP facilitates dialogue, debate and consensus around coverage and reimbursement and other policies that promote high priority research.

TABLE OF CONTENTS

EXECUTIVE SUMMARY	ix
PREFACE.....	xvi
INTRODUCTION	1
APPROACH AND SCOPE	3
RECOMMENDATIONS IN CONTEXT	10
REPORTING	
RECOMMENDATION 1. Reporting for Analytic Validity	13
CLINICAL VALIDITY	
RECOMMENDATION 2. Patient Population Selection for Assessing Clinical Validity.....	15
RECOMMENDATION 3. Appropriate Metrics for Clinical Validation.....	15
CLINICAL UTILITY	
RECOMMENDATION 4. Anticipating Clinical Pathways Related to Test Use	19
RECOMMENDATION 5. Outcome Measures for Clinical Utility	20
RECOMMENDATION 6. Use of Randomized Controlled Trials for Clinical Utility.....	21
RECOMMENDATION 7. Using Prospective-retrospective Clinical Trials.....	25
RECOMMENDATION 8. Using Single-arm Studies	28
RECOMMENDATION 9. Using Longitudinal Observational Studies.....	29
RECOMMENDATION 10. Using Decision-analytic Modeling Techniques.....	31
IMPLEMENTATION BARRIERS	
POSITION STATEMENT 1. Encouraging Public-private Collaborations	34
POSITION STATEMENT 2. Novel Policy Approaches to Promote Clinical Utility Studies	36
POSITION STATEMENT 3. Enhancing MDx Comprehension for Healthcare Professionals.....	38
References	39
Appendix A: Key Informants	49
Appendix B: Glossary of Terms.....	50

EXECUTIVE SUMMARY

The Center for Medical Technology Policy (CMTP) develops Effectiveness Guidance Documents (EGDs) to provide specific methodological recommendations targeted to clinical researchers and test developers regarding the design of clinical studies intended to inform decisions by payers, clinicians and patients. EGDs are envisioned to be analogous and complementary to Food and Drug Administration (FDA) guidance documents, but are focused on design elements that are particularly relevant to clinical and health policy decision making. The recommended methods aim to balance internal validity with relevance and feasibility, in order to provide these decision-makers with a reasonable level of confidence that the intervention improves net health outcomes.

A defined pathway for demonstrating how use of the test affects patient outcomes (clinical utility) is a critical unmet need for *molecular diagnostic (MDx) tests in oncology*, an umbrella term for any test that helps to identify inherited risk, diagnose or categorize cancer at the level of genes, proteins and their interactions within a cell. Focusing specifically on prognostic and predictive tests, there are a number of examples of MDx tests that are currently used in clinical practice to risk stratify cancer patients and target interventions, with accompanying evidence that use of these tests leads to improved health outcomes for patients. Nevertheless, there exists a large (and growing) group of potentially promising MDx tests that currently lack adequate evidence of clinical utility. Consequently, practice guideline committees and payers evaluating these tests often conclude that there is insufficient evidence to recommend clinical use or coverage, which correspondingly limits patient access. The purpose of this EGD is to close this gap between the presumed benefits of tests undergoing technology assessments and the information needs of payers, clinicians and patients. We accomplish this by first providing specific recommendations for designing studies to evaluate the clinical validity and utility of “actionable” MDx tests, those tests intended to guide clinical decision-making for patients with a known diagnosis of cancer.

The EGD’s overarching goal is to bring greater clarity and predictability regarding the evidence requirements of post-regulatory stakeholders. The benefits are two-fold: 1.) that test developers and researchers can anticipate reimbursement and clinical evidence expectations and plan studies appropriately, and 2.) that ultimately there is a common framework for evaluating the study results. In addition to developing methodological recommendations, the project team also generated position statements to acknowledge current barriers to evidence generation that go beyond better study designs and involve factors such as alignment of incentives, education, and the need for ongoing research collaborations. These statements reflect the aspirations of the project team and highlight key areas that need ongoing innovation for the field to realize the full potential of better evidence to support the clinical integration of MDx tests in oncology.

The EGD recommendations were created through an iterative process that incorporated the perspectives of major stakeholder groups, including researchers, clinicians, payers, industry, guideline developers and patient advocates. This stakeholder-driven process for developing the EGD provides an opportunity for input from potential EGD users and others who have a direct interest in how evidence is created and evaluated for actionable MDx tests in clinical oncology. This collaborative multi-

stakeholder process is essential not only for acceptance by researchers and test developers, but also to enhance the prospect that MDx tests evaluated by these standards (and receiving a positive assessment) will be used appropriately and reimbursed.

The project team recognizes that MDx tests, as well as evaluation methodologies for clinical validity and clinical utility, will continue to advance following the publication of this document. For this reason, we expect to revisit and update these recommendations periodically in light of new technical developments. We welcome feedback from all stakeholders regarding this version of the EGD; all comments received will be reviewed and considered for inclusion in future versions of this document. To submit feedback, please visit our website and use the email form provided on the “contact us” page: <http://www.cmtpnet.org/contact>

EFFECTIVENESS GUIDANCE DOCUMENT RECOMMENDATIONS

REPORTING

1. Test developers and researchers should ensure that the Analytic Validity (AV) of an MDx test has been established prior to the final assessment of clinical validity (CV). Following standard reporting guidelines will make the procedures used in establishing AV more transparent and more easily assessed for adequacy. Guidelines that could be used for this purpose include BRISQ, STARD, and AHRQ’s guidance on evaluation frameworks and genetic test evaluation.

CLINICAL VALIDITY

2. In planning clinical validity (CV) studies for MDx tests, developers must specify the patient population intended to benefit from the action or decision guided by the test result (the intended use population). For validation studies of all types (including, for example, the development of models), sufficient prior evidence from early validation studies must be obtained from the intended use population.
3. Clinical validation studies should report on the strength of an association between the MDx test and a specific disease state using metrics that are most useful to clinicians. The various sub-recommendations below attempt to highlight common validation issues that occur with single biomarker based tests, with “omics” based tests where the number of variables measured per patient may often be very high, as well as for both prognostic and predictive tests. However readers are encouraged to consult the references cited (in the full recommendation and in the accompanying Rationale section) for a detailed description of recommended methodological approaches.
 - i. When the test result and the clinical outcome are binary (e.g., presence or absence of a mutation and response to a targeted therapy), standard metrics for clinical validity include the test’s clinical sensitivity and specificity, and positive (PPV) and negative (NPV) predictive values for the disorder or outcomes.
 - ii. For test results that are continuous variables (e.g., a risk score), a clinical threshold or cutoff must be selected in order to generate a binary result (e.g. positive or negative). When the clinical outcome is binary (e.g. tumor response), receiver operating characteristic (ROC) curves can be useful to select optimal cutoffs that provide the performance needed (clinical sensitivity and specificity) for the biomarker or risk score underlying the test in order to support clinical decision-making for the specific intended use.

- iii. When the clinical outcome of interest is a continuous or time-to-event variable such as time to recurrence, regression methods may be used to model the relationship between the test result (discrete or continuous) and the outcome of interest. To classify patients into clinically actionable risk groups, it may be necessary to apply cutoffs to the results of the test and to the clinical outcome (e.g., disease-free survival at 5 years, tumor shrinkage of 50% or more). When applying a cutoff to a time-to-event variable, it is important to appreciate that a given test might more accurately predict early events than late events or vice versa.
- iv. For a predictive marker, an appropriate control group must be used to distinguish prognostic effects from predictive effects.
- v. To avoid ambiguity when reporting results, developers should use appropriate reporting standards and specifically define the terminology and concepts used.

CLINICAL UTILITY

4. To evaluate the clinical utility of an MDx test, the potential therapeutic actions or decisions (i.e, clinical pathways) that should be followed based on information obtained from the test must be specified in advance and must include all relevant treatment alternatives under consideration at the time of the test. These clinical pathways represent intermediate outcomes and the decision-making process guiding clinical pathway selection should be measured as part of the evaluation of clinical utility.
5. Studies to evaluate the clinical utility of MDx tests should include outcome measures that assess both potential benefits and harms of testing from the patient perspective, recognizing that these outcomes may occur at different time points and are the result of clinical management decisions guided by test results. Examples of typical outcome measures include clinical assessments of disease remission and progression, response to therapy, functional status as well as adverse events. Measures of benefits and harms should also routinely include patient-reported outcome measures, with the assurance that the selected measures are appropriate and validated for the clinical context. Clinical utility studies may also include important endpoints such as survival and downstream health care resource utilization; the decision to include these endpoints should be guided by the robustness of the existing evidence base regarding the specific clinical intervention prompted by the test result and its effects on relevant health outcomes. However process measures, such as changes in physician behavior, are typically insufficient to qualify as study endpoints, unless there exists a separate, robust body of credible evidence (as determined by widely accepted evidence review standards) linking clinical management decisions with relevant health outcomes. Studies designed to report intended care plans following an MDx test are insufficient for demonstrating clinical utility.

6. The clinical utility of an MDx biomarker should be assessed with randomized controlled trials that adequately evaluate the effectiveness of the clinical decision (treatment or other clinical pathway) relative to an appropriate control for both marker-positive and marker-negative patients. Enrichment designs which exclude from the study patients with a particular marker status should be avoided unless a clear and valid rationale exists for excluding non-marker status patients from study. Marker-based strategies which randomize patients to genomics-guided treatment vs. usual care partially duplicate the actions to be taken between the intervention and control arms, reduce statistical power, and therefore are not optimal because they require larger sample sizes to demonstrate an effect of the MDx test.
7. If an appropriately designed, powered and conducted clinical trial with banked biospecimens exists, then a properly conducted prospective-retrospective study is adequate evidence of clinical utility. Replication of study results (second study) and pooling of biospecimen samples from comparable RCTs are two approaches to address limitations related to causal inference and insufficient sample sizes.
8. Single-arm studies can be used to establish the clinical utility of an MDx test in cases where all of the following conditions are met: 1) the MDx test is being developed to be used with a drug that has already been FDA-approved on the basis of pivotal trials of a general population with regulatory endpoints such as survival or progression-free survival; 2) adequate archived tissue samples are not available to conduct a prospective-retrospective trial to assess clinical utility of the MDx test; 3) it is feasible to use response, variably defined as complete or overall response, as an endpoint in the single-arm study; and 4) there exists comparable response data in a non-contemporaneous comparative cohort.
9. Under limited, specified circumstances, longitudinal observational study designs such as prospective cohort studies, patient registries that explicitly include comparators, and multiple group, pretest/posttest designs (also called quasi-experimental, difference-in-difference design; regression discontinuity design) are acceptable options for assessing the clinical utility of MDx tests, provided that a compelling rationale for not doing an RCT is addressed (examples below), efforts to minimize confounding are documented, and good research practices for prospective observational studies are followed, including public registration of studies. Since the necessary parameters for evaluating the clinical utility of MDx tests (e.g., clinical characteristics of patients, test findings and interpretation, subsequent care and patient outcomes) are typically not found in secondary databases (including most electronic health records), the pursuit of retrospective observational studies is generally not adequate. Prospective observational studies may include the use of secondary databases as one component of the data collection effort, but must also include prospective data collection efforts to obtain the missing data or develop validated approaches to approximate these data elements from the existing secondary data.

10. Based on initial scenario modeling, formal decision-analytic modeling techniques can be used to elucidate the relationship between test results, corresponding clinical pathways and downstream patient outcomes in cases where an MDx test has established evidence of clinical validity and plausible evidence of clinical utility based on initial scenario modeling (a simplified approach to decision analysis that typically includes outcomes evaluated under 3 scenarios: base case, best case, worst case). Decision-analytic models are useful in the common situation where there is no direct evidence of clinical utility, as they provide explicit estimates of the likely effects of clinically validated test results on patient outcomes by linking separate sources of evidence, including quantifying the relationship between surrogate outcome measures and final patient outcomes. Models should include all patient-relevant benefits and harms related to the duration and quality of remaining life. Summary measures such as clinical events, life expectancy and quality-adjusted life years represent appropriate modeling outcome measures. Good modeling practices for diagnostic tests have been published and should be followed; these methods are labor and time intensive and are not recommended when there is a high degree of uncertainty about the underlying disease process, lack of a clinical intervention with known benefits, or when there is high uncertainty about the link between test results and the effectiveness of interventions.

EFFECTIVENESS GUIDANCE DOCUMENT POSITION STATEMENTS

1. We encourage sustainable new public-private collaborations in order to ensure the efficient implementation of clinical utility studies for specific MDx tests. The goal is to broaden the stakeholder engagement model that served as the foundation for the development of MDx evidence standards to larger groups of decision-makers. This facilitates a shared understanding of evidence thresholds as they apply to specific MDx tests, and ensures that individual clinical utility studies are maximally informative for clinical and coverage decision-making. In particular, we suggest that entities such as diagnostic and pharmaceutical companies, payers, cancer centers, accountable care organizations (ACOs), universities and professional societies collaborate to develop an interoperable research infrastructure and apply the EGD recommendations to individual study proposals and protocols. Over time, this will lower the barriers to developing evidence of clinical utility by providing consistent, predictable and uniform evidence standards for researchers and test developers, provide real-world opportunities for feedback and refinement of the standards, while providing sufficient flexibility for decision-makers to tailor their application to specific MDx tests.
2. We support the development and use of novel reimbursement policy approaches to promote clinical utility evidence generation for molecular diagnostic tests and other medical devices and drugs. Managed entry schemes encompass a broad range of policy tools that provide the flexibility to payers to cover innovative, emerging molecular diagnostic tests while generating valid evidence on the relative benefits and risks of these tests while they are used in clinical practice. Among the possible tools to be considered on a case-by-case basis are FDA-CMS parallel review and adaptive licensing (for companion diagnostics and in vitro diagnostic tests undergoing FDA review) and performance-based risk-sharing arrangements (potentially applicable to both LDTs and in vitro diagnostic tests undergoing FDA review), including the provision of coverage for patients in well-designed clinical trials to gather CU evidence for clinically promising MDx tests (coverage with evidence development).
3. We support initiatives that enable healthcare professionals to accurately interpret and communicate the results of molecular diagnostic testing to patients and their caregivers. Strategies include providing Continuing Medical Education credits (CME) for MDx-related training, as well as engaging professional societies to develop practice guidelines specifying the use and interpretation of MDx test results. In addition, test developers must work with both clinicians and patient advocates to design reporting templates that can be informative to all stakeholders, including patients. Therefore, we recommend that these groups collaborate to develop test reports that are maximally useful to patients, caregivers, and health care professionals.

PREFACE

The Center for Medical Technology Policy (CMTP) supports the development of Effectiveness Guidance Documents (EGDs) to provide specific recommendations on the design of prospective studies intended to inform decisions by patients, clinicians and payers. EGDs do not provide general methodological advice, but rather offer specific study design recommendations that are relevant to a defined clinical condition or category of clinical interventions. The purpose of EGDs is to better align the design of clinical research with the information needs of patients, clinicians, and payers. EGD recommendations will generally address one or more of the following elements of study design: patient inclusion/exclusion criteria, choice of comparators, research settings, selection of outcomes, duration of follow-up and other key elements of trial design that are most relevant to the topic of each guidance.

The primary audience for this EGD is test developers and clinical researchers who are developing research protocols for studies that are intended to be helpful to patients, clinicians and payers in making clinical or health policy decisions. This would include researchers from life sciences companies with clinical development responsibilities, or other clinical researchers receiving funding from public sources, foundations, etc. EGDs are intended to be analogous to FDA guidance documents in that they are also targeted to product developers and clinical researchers, and provide guidance on the design of clinical studies that are intended to support regulatory decision-making. EGD recommendations are not intended to establish standards for research to be considered adequate with respect to coverage, payment or pricing decisions. They are likely, however, to be aligned with the expressed evidence preference of public and private payers, as they are developed with payer input.

The methods recommendations in EGDs are guided by the objective of achieving an acceptable balance across a number of desirable dimensions, including validity, relevance, feasibility and timeliness. Overall, the objective of EGDs is to offer study design recommendations that would give decision-makers a reasonable level of confidence that the intervention studies would improve net health outcomes.

The recommendations in an EGD are influenced, and sometimes limited, by the available information. As new information about the treatment and management of a disease, or about the methods used to diagnose or otherwise test for disease status becomes available, the recommendations in an EGD may be modified.

There are a number of potential benefits of the creation and use of EGDs. First and foremost, they could help increase the consistency with which the body of clinical research reflects the information needs articulated by patients, clinicians and payers. In addition, EGDs could contribute to greater consistency of trial design across studies of related treatments within specific clinical conditions, allowing for higher quality meta-analysis and systematic reviews due to reduced heterogeneity across multiple studies. By considering existing regulatory guidance in the EGD process, it is hoped that EGDs will help to achieve optimal alignment between study design elements intended for regulatory approval and study design elements targeted to clinical and health policy decision-making.

There are three primary features that distinguish EGDs from the majority of other methods guidance documents. First, EGDs focus on a specific clinical area or category of interventions, while most other available methods guidance documents are more general and apply across a broad range of clinical conditions or technologies. Second, a number of the other documents provide guidance on reviewing the quality of existing studies, while EGDs provide recommendations for the design of future studies. Finally, we are not aware of any other documents that actively engage patients, clinicians and payers in the process of developing recommendations, with the goal of ensuring that the information needs of these decision-makers are given significant attention in generating methods recommendations.

PROCESS AND DEVELOPMENT OF CMTP EGDS

EGD recommendations are developed through an extensive consultative process involving a broad range of experts and stakeholders, including mechanisms for broad public review and comment. CMTP develops EGD recommendations with the support of a Technical Working Group (TWG) consisting of experts in clinical care and research methods specific to the clinical domain that is the focus of the EGD, and also including patient, clinician and payer representatives. Draft EGDs are made available for working group comments, and opportunities provided for input from the broader stakeholder community through one or more methods symposia to address the most complex or controversial issues. All feedback on the draft EGD is reviewed by CMTP staff as part of developing a “final” version of the EGD, which is posted on the CMTP website and widely distributed.

Full details about EGDs are available at <http://www.cmtynet.org/effectiveness-guidance-documents/>. The specific procedures used to develop this EGD are described in detail in the next section.

INTRODUCTION

Oncologists have long used a variety of tumor markers to aid in detecting, staging, and managing some cancers. For example, prostate specific antigen (PSA) has been used to screen for prostate cancer and CA125 is used to follow women during and after treatment for advanced epithelial ovarian cancer. The more recent development of utilizing a broad range of measurement technologies to test for DNA, RNA, proteins and metabolites in patients with cancer — in this document collectively referred to as molecular diagnostic tests — has the potential to transform oncology practice by emphasizing a molecular, rather than histologic approach to classification and management of various cancers. Molecular diagnostic (MDx) tests are now used to diagnose and stage certain cancers, provide information to help guide therapeutic selection and dosing, assess treatment response and aid in detection of residual or recurrent disease (McDermott et al., 2011).

Despite a burgeoning research enterprise in which a large number of potentially useful MDx biomarkers¹ have been identified, integration into clinical practice has been inefficient (Wilson et al., 2007). The reasons are multifactorial and range from technical issues involving assay variability and analytical validation (Marchio et al., 2011; Saijo, 2012; Simon, 2010) to problems with study design, interpretation and results reporting (McShane & Hayes, 2012; Wideroff et al., 2009), to challenges at the reimbursement, professional and consumer levels (Deverka et al., 2007). Nevertheless, oncologists and patients need useful MDx tests and there is currently no shared evidentiary framework for decision-making that can be applied to the available evidence by different stakeholders such as policy-makers, payers, clinicians and patients (Khoury et al., 2003). At the same time, the current market environment is characterized by relatively low barriers to entry and regulatory and reimbursement policies have not provided sufficient clarity for test developers to generate the evidence of net benefit to patients (clinical utility) needed for effective clinical use. Accordingly, the first three out of four evidence reviews of MDx tests currently in use conducted by the Evaluation of Genomic Applications in Practice and Prevention Working Group (EGAPP) resulted in judgments of “insufficient evidence to recommend for or against” (Khoury et al., 2010) and there is the widespread opinion that few MDx tests have met the standards needed to persuade the clinical community that they can be used to guide treatment decisions that will result in improved patient outcomes (Williams et al., 2012). The lack of evidence of clinical utility for MDx tests represents a serious stumbling block for developing clinical guidelines and ensuring access to testing through favorable coverage and reimbursement decisions (SACGHS, 2008).

This project aims to develop recommendations for improving the evidence base for MDx tests from the perspectives of end-users and policy decision-makers such as clinicians, patients and payers. While technology assessment organizations and guidelines development groups such as EGAPP, Blue Cross Blue Shield Association Technology Evaluation Center (BCBSA TEC), the National Cancer Coalition Network (NCCN), and the American Society of Clinical Oncology (ASCO) have frequently concluded that there is insufficient evidence to make recommendations supporting the use of recently introduced molecular diagnostics, these assessments often stop short of providing specific guidance regarding how researchers should design studies differently to overcome these deficiencies (Teutsch et al., 2009;

¹ Throughout this document, we have followed the convention of basic and clinical investigators and use the terms MDx, MDx test, marker and biomarker interchangeably (Poste et al., 2012).

BCBSA TEC, 2011; Engstrom et al., 2011; Robson et al., 2010). There have been several recent publications that provide recommendations regarding preferred study designs to evaluate the clinical utility of MDx tests under specific circumstances, such as when there are stored biospecimens available from previously conducted clinical trials or for an integral MDx test (Simon et al., 2009; Simon, 2010; IOM 2012a) While representing critical first steps from a clinical and statistical perspective, these standards need to be debated and elaborated by larger groups of stakeholders to ensure their relevance for clinical and policy decision-making. This guidance document therefore is unique in its attempt to use a stakeholder-driven approach to describe methodological recommendations for researchers and test developers to follow in their design of studies that meet the evidence needs of patients, clinicians and payers. Given significant public and private investments to improve our understanding of the molecular diversity of cancer and the early evidence that targeted approaches to treating cancer are an important clinical advance, we undertook this project based on the premise that MDx tests will only improve patient care if their integration into clinical practice is based on sound evidence of net benefit to patients.

APPROACH AND SCOPE

The procedure used to define the project scope and inform study design and planning was designed to balance the need for scientific innovation with appropriate evidence requirements for moving a biomarker into broad clinical use. The approach is distinctive because of its inclusive focus, collaborative process, emphasis on pragmatism, and goal of having the final set of recommendations strike an acceptable balance across a number of desirable dimensions, including internal validity, feasibility, relevance and timeliness.

This EGD was prepared through a multi-step process (see Figure 1). First, CMTP conducted a literature search to review major committee reports, evidence guidelines and synthesis documents, including MDx test evidence reviews performed by the EGAPP and the BCBSA TEC. Second, a semi-structured interview guide was developed, reflecting key findings from literature, technology assessments and systematic reviews. Third, initially through purposive sampling and subsequently employing snowball sampling, CMTP identified and interviewed experts representing all the major stakeholder groups, including researchers, clinicians, payers, industry, regulators, guideline developers, patients, and consumers. A total of 23 key informants were interviewed (Appendix A) and the project scope was initially defined during this process and then refined after additional consultation with a broader group of test developers and payers. The research plan was reviewed by an Institutional Review Board (IRB), which found that the plan met the criteria for exempt status.

The recommendations addressing the evidence gaps identified through the literature search and key informant interviews were developed through an ongoing engagement process with a technical working group (TWG) composed of clinical and methodological experts and a patient representative (see list p. iii) and a molecular diagnostic advisory group composed of industry experts (see list pp. v - vi). A TWG in-person meeting was held in Baltimore on September 16, 2011. Multiple teleconferences took place with the TWG between September 16 and May 23, during which time the primary framework for the recommendations was developed and revised. Molecular Diagnostics Advisory Group (MDAG) meetings with TWG participation were hosted by CMTP in Washington DC, on May 23, 2012 and in Baltimore on October 11, 2012. A series of working groups composed of TWG and MDAG members collaborated to address the discussion points from the meetings, further refining the scope, recommendations, and position statements. Figure 1 summarizes this interactive process.

CMTP maintains full authorship and editorial control over this EGD and all other materials related to this initiative. Authors of this EGD received no special compensation.

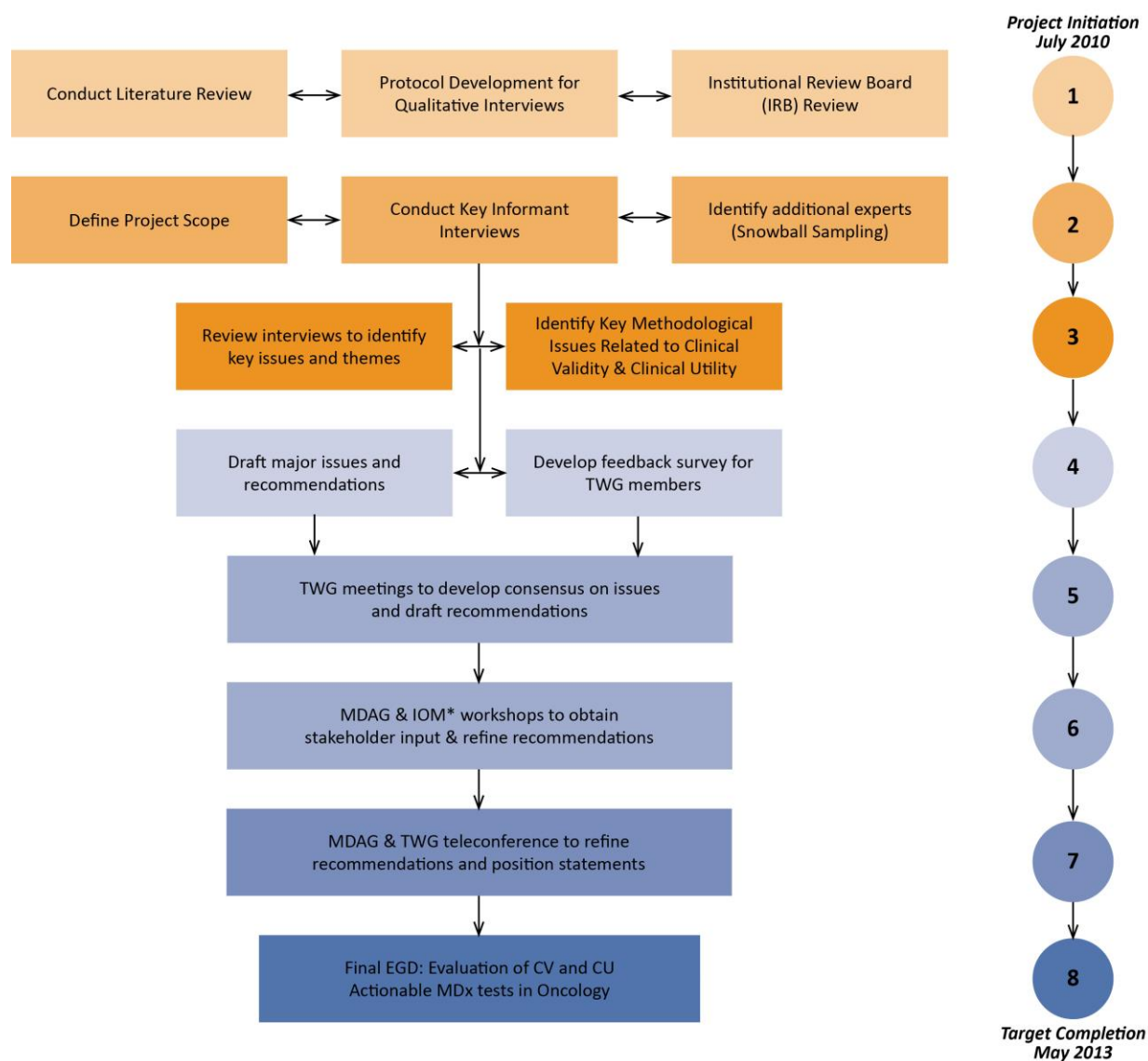


Figure 1: Process for Developing an Effectiveness Guidance Document (EGD) for Molecular Diagnostic Tests in Oncology

*Institute of Medicine. See first paragraph of page 10 for full description of process.

DEFINING THE PROJECT SCOPE

Recognizing that *molecular diagnostic tests in cancer* spans a very broad range of clinical applications (including screening, risk assessment, diagnostic, prognostic and predictive tests, as well as assays to monitor treatment response and tumor recurrence), a major emphasis of our key informant interviews was to have these experts prioritize the categories of tests they considered to have the greatest

clinical potential, yet biggest evidence gaps for informed clinical and reimbursement decision-making. The nearly unanimous opinion was that there was a great deal of promising research as well as both public and private investments directed at clinical translation for *predictive* and *prognostic* MDx tests, however these tests continue to face major stumbling blocks because of their lack of robust evidence of *clinical validity* and *clinical utility*. Additional details of how our project team further refined the project scope both as part of the key informant interviews and in consultation with our Technical Working Group at major milestones throughout the project are provided below.

TYPE OF TEST

Traditionally molecular tests for use in clinical practice in oncology have been categorized in the literature as either predictive or prognostic. However, this binary classification falls short for the purpose of defining evidence requirements, as some tests (e.g. Oncotype DX) can be both predictive and prognostic. A single biomarker may also be categorized differently across tumor types or stages of a tumor. Therefore, based on the advice of the TWG, the scope of our recommendations applies to **actionable tests**, meaning those that can lead to changes in the clinical management of patients (Dressler, 2012). Explicitly, the term **actionable MDx test** refers to tests that predict survival or other clinical endpoints independently of any specific treatment ('prognostic test'), tests that predict response to treatment, ('therapy guiding' or 'predictive test') tests that assess response to treatment ('monitoring test') and tests that are used to identify the risk of organ-based toxicities or altered metabolism and/or response to cancer drugs ('pharmacogenomic test'), **as long as the test result leads to some type of clinical intervention**. The TWG also suggested that since these tests were guiding patient care decisions and cancer was a potentially life-threatening clinical condition, all of the tests should be classified as "high-risk" tests when characterizing the potential benefits and harms to patients. This classification is analogous to the Federal Code of Regulations definition of a significant risk medical device (e.g., MDx test) as a test "...for a use of substantial importance in diagnosing, curing, mitigating, or treating disease, or otherwise preventing impairment of human health and presents a potential for serious risk to the health, safety, or welfare of a subject" (FDA, 2011a).

Recognizing that in the U.S. there are two regulatory paths for MDx tests to reach the market, either as a laboratory developed test (LDT) where the laboratories are regulated by the Centers for Medicare and Medicaid Services under the Clinical Laboratory Improvement Amendments (CLIA) or as an *in vitro* diagnostic device regulated by the FDA, our group debated the merits of including companion diagnostics tests as within the scope of our EGD recommendations. Unlike LDTs, *in vitro* diagnostic (IVD) devices developed to guide clinical decision-making for a specific drug, widely termed "companion diagnostics," are subject to FDA review. In a recently published draft guidance, the FDA defined an IVD companion diagnostic device as "an *in vitro* diagnostic device that provides information that is essential for the safe and effective use of a corresponding therapeutic product" (FDA, 2011c). Because the results of companion diagnostic tests guide important treatment decisions, the FDA regulates this category of MDx tests as high risk products. The FDA expects that ideally, "a therapeutic product and its corresponding IVD companion diagnostic device would be developed contemporaneously, with the clinical performance and clinical significance of the IVD companion diagnostic device established using data from the clinical development program of the corresponding therapeutic product." Under these circumstances, the "locked down" (analytical and clinical/biological validation established) companion test would be used in the registration trials for the drug, thus

generating evidence of clinical utility for the drug-device combination. Evidence of clinical validity may be inherent if the drug and companion test targets are the same, or may be available from preclinical data. In cases where the companion diagnostic test is developed separately from the underlying therapy (e.g., a diagnostic test is intended to serve as a companion to an already-approved therapy), the FDA would require approval or clearance of the test as a device, with evidence of its clinical significance in connection with the therapy (FDA, 2011c).

Although only a few real-world examples are as yet available and the final FDA guidance has not been published, our project team decided it would be instructive to analyze the Blue Cross Blue Shield Association (BCBSA) technology assessment of the case of the BRAF gene mutation testing to select patients with melanoma for treatment with BRAF kinase inhibitors (BCBSA TEC, 2011). The case findings were discussed with our TWG (one member represented BCBSA TEC) in light of FDA's approval criteria for all recently approved companion diagnostics in oncology (e.g., vemurafanib). In the interest of priority-setting, at the time of this writing the authors of this EGD take the position that the FDA approval process for companion diagnostics is likely to result in *sufficient evidence of clinical utility for the diagnostic-drug combination*. Accordingly, no methodological recommendations are provided specifically for the design of clinical utility studies for co-developed MDx tests that are approved by the FDA, although we do provide guidance for situations where an MDx test is developed at a later time point for an already-approved drug (Recommendation 8).

MDx tests are sometimes also further sub-divided based on the particular technology platform that is used (first generation techniques such as Sanger sequencing) or next generation sequencing technologies that enable accurate characterization of large volumes of sequence information. After discussing the applicability of the EGD recommendations to the full spectrum of MDx tests, including tests to both enumerate and analyze circulating tumor cells, the group decided that the recommendations could be generalized across all test types. Therefore, the EGD recommendations should be considered to be "technology platform agnostic"; the focus is on considerations related to establishing how use of the test identifies clinically relevant phenotypes and improves health outcomes.

TUMOR TYPES

Malignancies in oncology take many forms, with abnormal cell proliferation originating in epithelial cells, hematopoietic cells, connective tissue and germ cells. Within these cell types of origin, extensive sub-categorization exists. The promise of genomics is to move beyond histological classification of tumor subtypes to a molecular-based understanding of tumor biology. For the purposes of this EGD, the goal was to gain agreement regarding the broadest category of cancer types for which a single set of meaningful recommendations could be developed in order to have maximum impact. The key informants interviewed for this project initially framed the scope in terms of two broad categories of malignancies: solid tumors and hematologic malignancies. Their rationale was based on concerns that methodological considerations between hematologic and solid tumors were potentially too varied for all tumor types to be considered within one EGD. However, as the engagement process continued, TWG members and advisors later suggested revising the scope, expressing confidence that a useful set of recommendations could be generalized across the spectrum of both solid and hematologic malignancies. We also discussed whether the EGD recommendations might apply to pediatric as well

as adult cancers. While not necessarily having a biologic rationale for our decision, we concluded that because our TWG did not include a pediatric oncologist or pediatric patient advocate, and having not considered this distinction from the outset, we would limit the scope of our recommendations to the use of MDx tests in adult oncology.

RECOMMENDATIONS – PRIORITIZING THE EVIDENCE GAPS

The **ACCE** framework (Analytic Validity, Clinical Validity, Clinical Utility and Ethical, Legal and Social Implications) provides a model process for collecting, evaluating, interpreting, and reporting data about genetic testing (Figure 2; Haddow & Palomaki, 2003). This framework is widely recognized and directly applicable to molecular diagnostics in oncology and was therefore used to conceptualize evidence requirements for MDx tests in this EGD. Although it is often said that “clinical utility” is a poorly understood term, our literature review revealed good consensus regarding how the term is defined outside of the regulatory context by researchers, clinical guidelines groups and payers to mean whether use of the test leads to improved patient outcomes compared to an alternative (Poste et al., 2012; Williams et al., 2012; Khoury et al., 2009; Rogowski et al., 2009). Similarly analytic validity addresses whether the MDx provides accurate information and clinical validity assesses how well the test result correlates with a clinical outcome (refer to Glossary for complete definitions). Based on this framework, our key informants described the greatest unmet need for better evidence was in the area of clinical utility, but agreed there were common problems that occurred with clinical validity studies. Since clinical validity should always be established before proceeding to an evaluation of the clinical utility of an MDx test, they recommended that we address certain critical aspects of clinical validity as well.

This advice aligned with the findings from our literature review which emphasized the lack of clinical utility data (Khoury et al., 2003) and flaws in many of the published clinical validity studies (McShane, 2012). Studies of analytic validity are rarely published (Veenstra et al., 2013) preventing an assessment of study limitations and current systems of regulatory oversight focus their efforts on assuring the technical efficacy of tests. The Centers for Medicare and Medicaid Services (CMS) oversee analytic validity in clinical laboratories under CLIA and the FDA has developed a guidance document on ensuring analytic validity for tests submitted to them for review and approval (FDA, 2007). Additionally, a number of professional groups are active in addressing standards for ensuring analytic validity (Jennings et al., 2009; CDC, 2009) and professional societies such as the College of American Pathology (CAP) oversee voluntary accreditation and sample exchange programs that specifically address analytic validity. Therefore although ensuring analytic validity is critical, recommendations to design analytic validity studies were not viewed to be a priority at this time by our group.

The TWG and other stakeholders discussed whether recommendations were needed for studies of cost-effectiveness, or evidence of value. While recognizing the growing importance of this type of evidence for some payers, the group decided to keep the primary focus on the standards for the generation of clinical utility, since this was the area of endeavor most in need of expert methodological guidance according to our review. However, consistent with the ACCE definition of clinical utility as “the balance of benefits and harms associated with the use of the test in practice, including improvement in measurable clinical outcomes and the usefulness or added value in decision-making compared with not using the test” (Haddow & Palomaki, 2003), the group did agree that a

comprehensive evaluation of the clinical utility of an MDx test should include measures of how use of the test affects health care resource utilization. The corresponding assignment of specific costs and formal cost-effectiveness analysis were considered to be out of scope for this EGD.

There are other frameworks commonly used for structuring evidence reviews of diagnostic tests, such as the **PICO** typology (**P**atient population, **I**ntervention, **C**omparator and **O**utcomes) that have been shown to play an important role for defining the question of interest (Samson & Schoelles, 2012; Veenstra et al., 2013). Therefore throughout the process we emphasized specific aspects of the PICO typology where appropriate. Also, during our project other frameworks have been published that are also useful for organizing the evidence requirements and corresponding methodological recommendations. We have modified our approach to reflect these enhancements and ensure that the EGD is up-to-date when published (see next section).

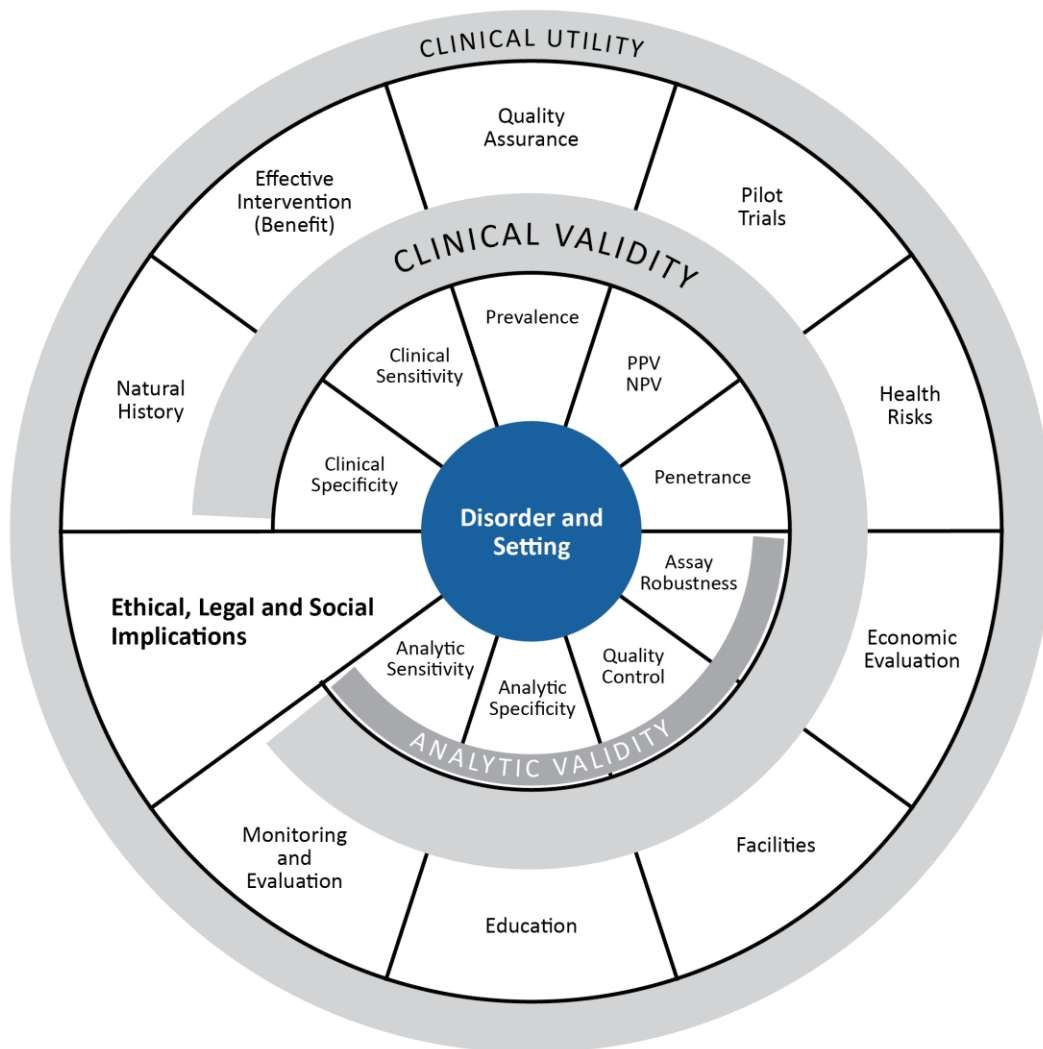


Figure 2: The ACCE Evaluation Process for Genetic Testing

Table 1. Scope of MDx EGD	
IN	OUT
Type of Test	
Actionable	Companion diagnostics
High risk	Low/Moderate risk
New test/existing drug	
Stand-alone test	
First generation assays, Next gen assays, Circulating tumor cells	
Target Condition	
Solid tumors	
Hematologic malignancies	
Adult patients	Pediatric patients
Recommendations	
Clinical validity	Analytic validity
Clinical utility	Formal cost-effectiveness-analysis
Implementation barriers	

RECOMMENDATIONS IN CONTEXT

Over the past several years, *The Roundtable on Translating Genomics-Based Research for Health* of the Institute of Medicine (IOM) has been actively involved in efforts to clarify the key issues in the evaluation of molecular diagnostic tests, including discussion of the standards and methods for evidence generation, barriers to evidence generation, and policy options for improving evidence generation, oversight and reimbursement of molecular diagnostic tests in the translation to clinical practice. For example, on November 17, 2010, the IOM hosted a workshop called “Generating Evidence for Genomic Diagnostic Test Development” (IOM, 2011). A year later, on November 15, 2011, the IOM hosted another workshop called “Facilitating Development and Utilization of Genome-Based Diagnostic Technologies” (IOM, 2011). By design, neither of these workshops was intended to develop specific recommendations. However, as noted below, many of the topics discussed in these roundtables are also topics addressed by the CMTP recommendations presented below and serve as useful documents to frame the issues as well as help prioritize evidence gaps and potential solutions. On May 24, 2012, CMTP co-hosted a workshop with the IOM to specifically address the “Evidence for Clinical Utility of Molecular Diagnostics in Oncology” with an interactive agenda focused on discussing specific study methodologies as well as innovative models for developing better evidence that would meet the needs of all stakeholders (IOM, in press).

In addition, an IOM committee was convened in 2010 to investigate and recommend sound principles for appropriate development and translation of “omics-based” tests (genomics, proteomics, transcriptomics, metabolomics, epigenomics, etc.) from research laboratories into clinical trials (Micheel et al., 2012). This charge was given to the committee specifically in response to a series of ongoing clinical trials in which genomics-based predictive tests of questionable utility were being employed to direct the treatment of cancer patients. As discussed below, the committee report released in early 2012 does include specific recommendations for development and evaluation of “omics” tests.

In this sense, the CMTP recommendations are more aligned in purpose with the 2012 “Evolution of Translational Omics” report than the other meeting summaries noted above. The Omics report is methodologically focused and discusses test development in three stages: discovery, validation, and clinical utility. The main overlap with the CMTP recommendations occurs in the second stage, where the test has already been validated and the investigator is planning a study to assess the clinical utility of the test, with multiple design options depending on the intended clinical use of the test and availability of biospecimens from previous clinical trials. However, the CMTP EGD differs from the Omics report in that it is technologically agnostic, does not address issues related to discovery or analytic validation (out of scope) and is intended to address the information needs of payers, clinicians and payers, not primarily regulators. Correspondingly, the CMTP EGD focuses on specific study design recommendations (including studies other than RCTs) as well as position statements to acknowledge the current barriers to evidence generation that will require stakeholder-driven partnerships and policy solutions in addition to better study designs.

Since the start of this project, the Methodology Committee of the Patient-Centered Outcomes Research Institute (PCORI) has published a Methodology Report (Helfand et al., 2012) that includes standards for designing studies of diagnostic tests based on expert review of the literature (Carlos et

al., 2012), followed by a public comment period where the standards could be modified. While they specifically stated that they excluded “genetic, genomic or pharmacogenetic/omic testing” as beyond the scope of this initial report, methodological reasons were not provided for the exclusions. As the final standards are quite general, they are not in conflict with any of the recommendations described in this EGD and if anything, provide additional support and rationale for the development of specific recommendations in the current context of molecular diagnostics and cancer.

There have also been several publications describing at a high level acceptable study designs for clinical validity and clinical utility studies, but these are typically in the context of integral assays (biomarkers used to determine eligibility, assign treatment or assess outcome) that have also emphasized corresponding regulatory requirements, such as the need for an investigational device exemption from the FDA and the need to conduct the assay in a CLIA-accredited laboratory (Schilsky et al., 2012). This EGD recognizes that regulatory compliance is critical, therefore we have reviewed this literature and attempted whenever possible to align our recommendations with those of researchers and guidelines committees, in addition to FDA requirements (Schilsky et al., 2012; Febbo et al., 2011).

Finally, while the ACCE framework is extremely useful for structuring the evaluation of evidence for genomic tests, there have recently been refinements to the model, including the suggestion that MDx test development be conceptualized in a comparable manner to the phases of drug development (Poste et al., 2012). This enhancement was recently expanded to include six phases in the development of genomic tests, adding a “Phase 0” that includes biomarker discovery and divides “Phase 4” (corresponding to post-marketing surveillance in drug development) into two separate phases that account for both comparative effectiveness research and assessment of population impact (Lin et al., 2012). The advantages of this framework is that it emphasizes the cumulative, “ground-up” aspect of evidence development and allows evidence review groups with a refined tool for communicating where evidence for a particular test is absent and what types of research is needed in a general sense. However the framework does not specify any evidence thresholds for decision-making nor specify any hierarchy of study designs – that is a major difference with our EGD. Nevertheless, we felt the framework was a useful tool for organizing our recommendations and indicating how they relate to other related efforts to improve the quality and availability of evidence for MDx tests. Therefore we contacted the authors and received permission to adapt their framework in order to more effectively communicate the intended purpose of our recommendations to relevant stakeholders (see Figure 3).

We describe four major clinical phases covering analytic validity, clinical validity and clinical utility, recognizing that biomarker discovery and early assay development activities occur in Phase 0 (not shown) and the assessment of population impacts, including assessments of cost-effectiveness (Phase 5, also not shown) go beyond the scope of our current EGD. Our chart differs from the original depiction of the framework only in the sense that we describe specific study design or reporting standards at each of the four phases of evidence development. Our intention was to show how our specific recommendations fit within this useful, but more general model for the particular case of actionable MDx tests in oncology. This will enable the larger community of MDx stakeholders to provide feedback regarding the relevance and practicality of our recommendations for clinical and policy decision-making, using a common framework for the discussion.

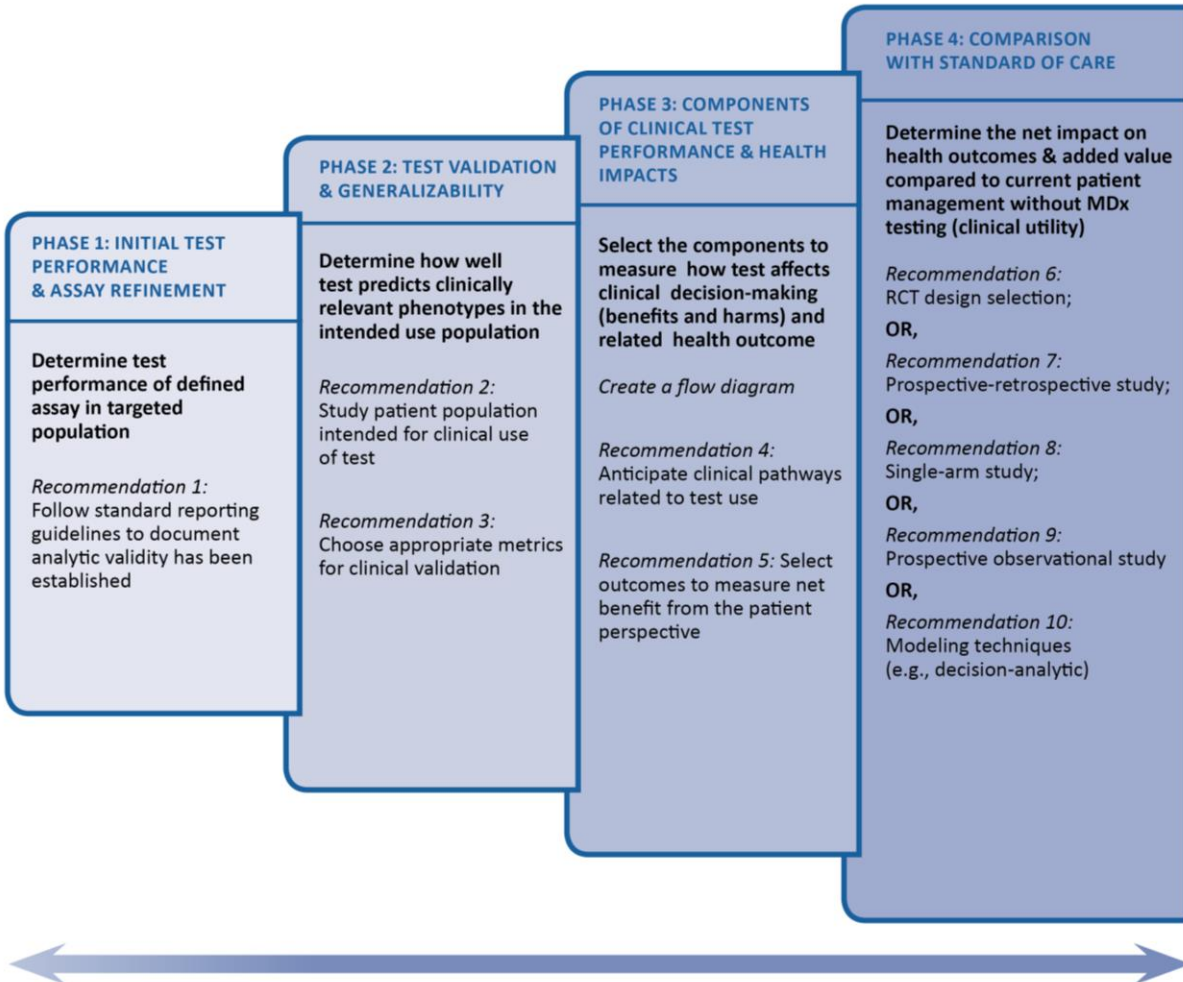


Figure 3: Clinical Phases of Test Development

RECOMMENDATIONS

REPORTING RECOMMENDATION

The generation of practical, high quality evidence of clinical validity and clinical utility depends on the assumption that studies are being conducted with an analytically validated MDx test. Ensuring the technical performance of an assay (e.g, analytic accuracy, precision and reproducibility) is essential before embarking on studies of clinical validity for actionable MDx tests in oncology, both from a patient and a regulatory perspective. There are several published guidelines for the reporting of study results that serve as useful “checklists” to ensure that researchers and test developers have paid adequate attention to the inter-related issues of preanalytic factors and analytic validity before proceeding with subsequent clinical validation studies.

RECOMMENDATION 1: Reporting for Analytic Validity

Test developers and researchers should ensure that the Analytic Validity (AV) of an MDx test has been established prior to the final assessment of clinical validity (CV). Following standard reporting guidelines will make the procedures used in establishing AV more transparent and more easily assessed for adequacy. Guidelines that could be used for this purpose include BRISQ, STARD, and AHRQ’s guidance on evaluation frameworks and genetic test evaluation.

Rationale: Although specific methodological recommendations related to analytical validity (AV) were excluded from the scope of this guidance document, ensuring AV prior to the final assessment of clinical validity (CV) is critical to improving the evidence base for molecular diagnostics in oncology. We recognize that in the early phases of biomarker discovery, the technical specifications of the assay may not be the same as those defined for later clinical validation studies. However once the intended use for test is determined, it is imperative that proper methods for establishing AV be followed (IOM 2012a). In CV studies where AV has not been properly established, it may be unclear whether suboptimal results are attributable to the CV study itself or inadequate AV. Several existing sets of standards provide guidance to researchers with regard to the conduct and reporting of preanalytic and analytic processes (Betsou et al., 2010; Moore et al., 2011; Bossuyt et al., 2003; Altman et al., 2012). To strengthen communication regarding biospecimen-related research, the Biospecimen Reporting for Improved Study Quality (BRISQ) describes reporting elements related to the collection, processing and storage of human biospecimens. BRISQ elements are prioritized into 3 tiers based on relative importance with the first tier representing items the committee deemed most critical to report (Moore et al., 2011). The Standards for Reporting of Diagnostic Accuracy (STARD) offers a list of guidelines for reporting each component of studies of diagnostic accuracy

(Bossuyt et al., 2003). Finally, the Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK) checklist recommends information that should be reported in all published tumor marker prognostic studies, including preanalytic procedures such as the preservation and storage of biospecimens, as well as analytic issues such as assay methods and quality control (Altman et al., 2012). Despite the fact that the original scope of the REMARK recommendations was predominately focused on prognostic studies, the guidelines are relevant to predictive studies and studies evaluating multivariable classification indices (McShane & Hayes, 2012). The Agency for Healthcare Research and Quality recently published a checklist based on a review of all the published reporting guidelines as well as in consultation with diagnostics experts. While intended for systematic reviews, the items represent a useful guide to the types of factors important to consider in ensuring that a test has adequate evidence of analytic validity (Sun et al., 2011).

METHODOLOGICAL RECOMMENDATIONS ON CLINICAL VALIDITY

To assess clinical validity of a molecular diagnostic test, it must be demonstrated that there is an association between the test result and the clinical condition of interest. In the setting of actionable MDx tests in oncology, examples of relevant clinical outcomes include measures of disease prognosis and treatment response, such as overall survival, disease-free survival or some other clinically meaningful metrics of response to therapy such as shorter duration of symptoms (Poste et al., 2012). When one must wait a period of time to observe an outcome of interest, it can be tempting to rely on intermediate outcomes, but this practice can be problematic if the intermediate outcomes do not reliably predict the definitive clinical endpoint of interest (Prowell & Pazdur, 2012). Another common flaw in clinical validity studies is that early validation work is conducted in highly selected populations that are not representative of how the test will be used in clinical practice, yet the results are often extrapolated from these initial studies without subsequent validation in the appropriate intended use population. In the past, published biomarker validation studies have been characterized by significant design and analysis limitations, however best practices have been described for both prognostic and predictive markers (Royston et al., 2009; Altman et al., 2009; McShane & Hayes, 2012; Janes et al., 2011). These best practices emphasize issues such as attention to data quality, sample size, patient population, choice of outcome measures and appropriate statistical analysis and results interpretation. Our goal here is not to replicate these efforts, but to highlight those methodological standards related to clinical validity felt to be most critical for informed decision-making by a group of stakeholders including payers, clinicians, patients, test developers, researchers and research funders.

RECOMMENDATION 2: Patient Population Selection for Assessing Clinical Validity

In planning clinical validity (CV) studies for MDx tests, developers must specify the patient population intended to benefit from the action or decision guided by the test result (the intended use population). For validation studies of all types (including, for example, the development of models), sufficient prior evidence from early validation studies must be obtained from the intended use population.

Rationale: Early validation studies for MDx tests, especially for biomarker-based prognostic tests but often for predictive tests as well, are often conducted using a convenience sample of available specimens (Simon, 2008; Woodcock 2010). It is well recognized that many MDx tests have been developed using tumor samples that have been previously collected and stored. However, the patient groups represented in these studies is typically too varied (by treatment, stage, and standard prognostic factors) to produce generalizable information. Preliminary and exploratory studies early in test development might validly use less representative patient populations, but efforts should be made to identify a specific intended use for the MDx test as early as possible in the development process. As test development proceeds, an unbiased clinical validation of a MDx test should ensure that the test sets used for validation should be drawn from the intended use population and be independent of any training data sets used to develop the test (Simon, 2005a). Subsequent external evaluation of patient benefit associated with the MDx test's use in clinical decision-making (clinical utility) should likewise draw from populations representative of the group(s) likely to be treated in clinical practice (Simon, 2005b). Reporting tools such as STARD (Bossuyt et al., 2003) and REMARK (Altman et al., 2012) specify procedures for reporting 1) how CV study patients were selected, and 2) whether the patient population selected for clinical studies is representative of patients receiving the test in clinical practice (see also Reitsma et al., 2009; Whiting et al., 2011). Use of these reporting guidelines will improve the quality and transparency of clinical validity studies.

RECOMMENDATION 3: Appropriate Metrics for Clinical Validation

Clinical validation studies should report on the strength of an association between the MDx test and a specific disease state using metrics that are most useful to clinicians. The various sub-recommendations below attempt to highlight common validation issues that occur with single biomarker based tests, with "omics" based tests where the number of variables measured per patient may often be very high, as well as for both

prognostic and predictive tests. However readers are encouraged to consult the references cited here and in the Rationale section for a detailed description of recommended methodological approaches. (See McShane & Polley, 2013; Polley et al., 2013; Altman et al., 2009; Janes et al., 2011; Moons et al., 2009a; Moons et al., 2009b; Royston et al., 2008; Royston et al., 2009; Wittes 2008.)

- i. When the test result and the clinical outcome are binary (e.g., presence or absence of a mutation and response to a targeted therapy), standard metrics for clinical validity include the test's clinical sensitivity and specificity, and positive (PPV) and negative (NPV) predictive values for the disorder or outcomes.**
- ii. For test results that are continuous variables (e.g., a risk score), a clinical threshold or cutoff must be selected in order to generate a binary result (e.g. positive or negative). When the clinical outcome is binary (e.g, tumor response), receiver operating characteristic (ROC) curves can be useful to select optimal cutoffs that provide the performance needed (clinical sensitivity and specificity) for the biomarker or risk score underlying the test in order to support clinical decision-making for the specific intended use.**
- iii. When the clinical outcome of interest is a continuous or time-to-event variable such as time to recurrence, regression methods may be used to model the relationship between the test result (discrete or continuous) and the outcome of interest. To classify patients into clinically actionable risk groups, it may be necessary to apply cutoffs to the results of the test and to the clinical outcome (e.g., disease-free survival at 5 years, tumor shrinkage of 50% or more). When applying a cutoff to a time-to-event variable, it is important to appreciate that a given test might more accurately predict early events than late events or vice versa.**
- iv. For a predictive marker, an appropriate control group must be used to distinguish prognostic effects from predictive effects.**
- v. To avoid ambiguity when reporting results, developers should use appropriate reporting standards and specifically define the terminology and concepts used.**

Rationale: In establishing the clinical validity of an MDx test, developers should demonstrate the association between the test and the specific disease using metrics that will be most relevant to clinical decision-makers. When the clinical disease state is binary, preferred metrics are clinical sensitivity, clinical specificity, positive-predictive value (PPV), and negative-predictive value (NPV). Knowledge of the disease prevalence in the population to be tested is required to compute PPV and NPV. Estimates of sensitivity and specificity should be provided with measures of uncertainty such as 95% confidence intervals. How much weight to give to false positives versus false negatives depends on the clinical context. For example, when a biomarker is tested to determine applicability

of a treatment, and the patient has no other treatment option, a high NPV is crucial to ensure that patients who will not benefit from this treatment are identified with a very low error rate. However, if more and possibly better treatment options are available, a higher PPV would be desirable for discriminating patients who will truly benefit. The optimum cut-point for clinical decision-making can be selected using an ROC curve to plot clinical sensitivity and specificity pairs associated with levels of the MDx biomarker (Pepe et al. 2007; Pepe et al., 2008). However, the area under the ROC curve should not be used as the only metric to compare or evaluate clinical validity.

Prognostic biomarkers are typically evaluated as part of a multivariate analysis to build a model for predicting a particular outcome. Methods to develop and analyze multivariate models have been described in detail (Royston et al., 2009; Moons et al., 2009a), although there is no consensus on a uniformly best method for all applications. Prognostic biomarkers are best examined in a prospective cohort study (Moons et al., 2009b) or possibly in the control arm of a randomized clinical trial. The preferred study design for validating a predictive biomarker is an RCT that compares two treatments where biomarker status is available for all patients at baseline (not an enrichment design). When the predictive biomarker is a continuous measure, a useful approach for choosing a cutoff is to examine marker by treatment predictiveness curves (Janes et al., 2011). Predictiveness curves display a clinical outcome (e.g., 5-year disease-free survival rate) as a function of biomarker value, separately for each treatment arm, but plotted together. This allows one to assess which treatment yields greater benefit at each biomarker value and to estimate the proportion of patients who will benefit from each treatment.

To encourage transparent and complete reporting of study design and statistical analyses and promote reproducibility, reporting of test validation studies should utilize appropriate standards such as REMARK (Altman et al., 2012).

METHODOLOGICAL RECOMMENDATIONS ON CLINICAL UTILITY

Given that the intended clinical use of this particular subset of “actionable” MDx tests is to inform clinical decision-making for patients with a known diagnosis of cancer, all decisions in this context are considered “high stakes” choices; incorrect decisions can have immediate and serious impact on the patient’s clinical course, health-related quality-of-life, and survival. **Tests for these purposes can therefore be considered high-risk medical decision tools and accordingly high standards of evidence apply.** For this reason, in general, randomized controlled trials (RCTs) are assumed to be the preferred methodology for assessment of clinical utility of actionable MDx tests in oncology, and the clinical validity of the test product will be “fully specified and locked down” prior to initiating prospective evaluations of clinical utility (Micheel et al., 2012). Recent publications have suggested that studies undertaken to demonstrate clinical utility of significant risk MDx tests must require that these tests be

performed in a CLIA-accredited laboratory and developers should consult with the FDA regarding the need for an investigational device exemption or application (Schilsky et al., 2012).

In the earliest stages of assay development, there should be a systematic approach to planning for the evidence-based translation of MDx tests into clinical practice. This approach should begin with describing the proposed clinical utility of the test in terms of a flow diagram (Figure 4) that outlines at a conceptual level the intended clinical use in practice and the associated primary patient outcomes, analogous to defining the primary study objectives for a clinical trial (Lord et al., 2009).

Recommendations 4 and 5 provide more detail regarding how this should be done; a process that supports the depiction of the flow diagram as a hypothetical RCT, showing key elements such as the test target population, existing test strategies, treatment alternatives and the pathway linking testing to patient outcomes. This hypothetical RCT specifies the type of comparative evidence that is needed to measure the proposed differences between the new test and usual care. In most cases, an RCT will need to be conducted. Recommendations for appropriate RCT designs are provided in Recommendation 6. Specific circumstances that allow alternative study designs, including prospective observational studies are described in Recommendations 7, 8 and 9.

There may be situations when prospective studies may not need to be conducted, as there are existing sources of evidence (published studies, observational data) that can be used to construct a “chain of evidence” that makes the link between testing and patient outcomes, (i.e., between clinical validity and clinical utility) based on the use of modeling (Recommendation 10) (Trikalinos et al., 2012; Plevritis, 2005). The value of developing the flow diagram early in the process of planning for the proposed clinical utility of the test is two-fold: 1) it helps the researcher decide whether a prospective study is necessary by identifying existing data sources that help establish the link between test result and patient outcome(s), and 2) it identifies the critical missing data elements, thereby helping the researcher to design more efficient studies. (See Figure 4, which provides a schematic example of a flow diagram to assess known information for a clinical decision-making setting. Red boxes indicate decision steps for which information does not exist or is inadequate. For illustration purposes, the figure includes some hypothetical data and reference sources for each pathway. An actual flow diagram would specify the information available and sources for each branch in the diagram to provide a more detailed map of the type of information still needed to develop the test fully.)

At the highest level, this approach is comparable to the development of an analytic framework by EGAPP and other evidence-based review groups. The analytic framework makes explicit the series of steps linking the test result to the associated clinical interventions and patient outcomes. Key questions posed to review groups are correlated with intermediate steps in the analytic framework and data from a variety of reputable sources is used to create a chain of evidence where direct evidence is lacking regarding the effect of the test on patient outcomes (Teutsch et al., 2009; Pettiti et al., 2009). However analytic frameworks lack the specificity and detail of the flow diagram described above, which can be thought of as a *decision tree*, depicting multiple alternative uses of the MDx test and its comparators and the specific potential health outcome consequences of altered decisions (Samson & Schoelles, 2012).

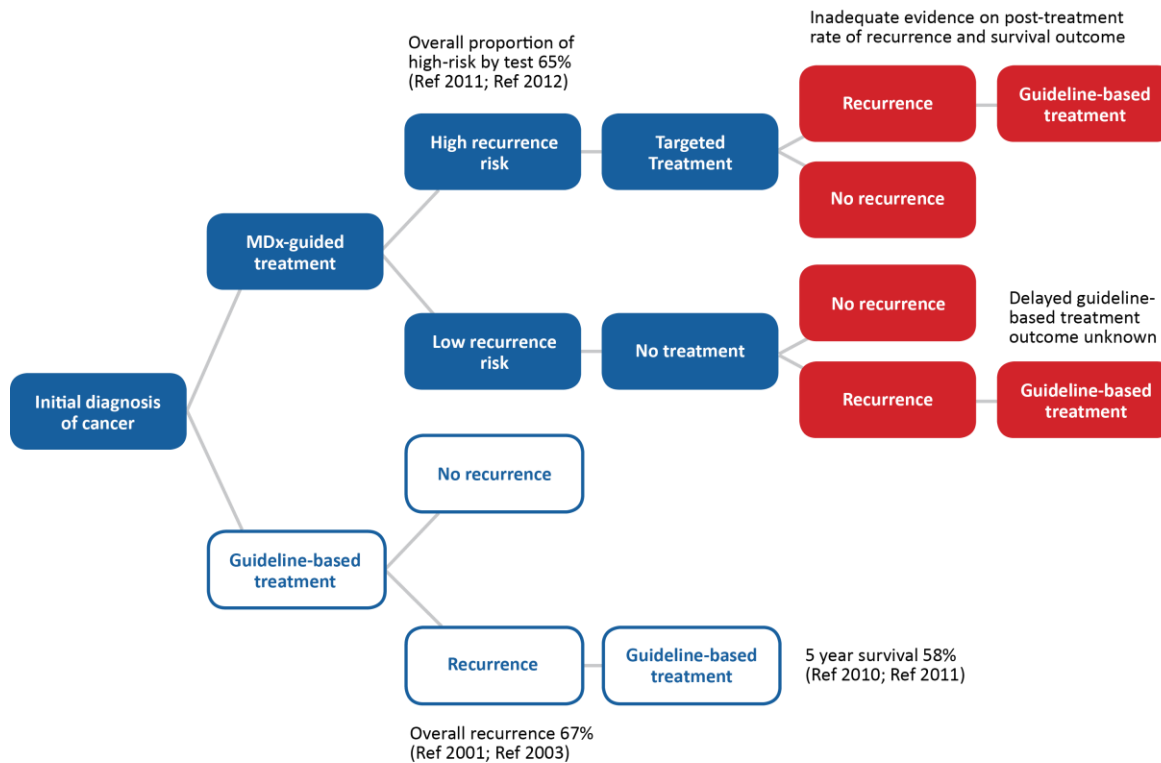


Figure 4: Example Flow Diagram

RECOMMENDATION 4: Anticipating Clinical Pathways Related to Test Use

To evaluate the clinical utility of an MDx test, the potential therapeutic actions or decisions (i.e, clinical pathways) that should be followed based on information obtained from the test must be specified in advance and must include all relevant treatment alternatives under consideration at the time of the test. These clinical pathways represent intermediate outcomes and the decision-making process guiding clinical pathway selection should be measured as part of the evaluation of clinical utility.

Rationale: MDx biomarkers are being discovered rapidly, yet their use in clinical practice is relatively limited (Febbo et al., 2011; Rogowski et al., 2009; Pusztai et al., 2003). Prognostic biomarkers have been particularly slow to translate into clinical practice because of significant deficiencies in study design methods and statistical analyses,

including the absence of a clinically meaningful marker question, i.e., the studies were designed without pre-specifying the clinical application (McShane & Hayes, 2012). One notable exception is in the area of hematologic oncology where molecular or minimal residual disease (MRD) monitoring is routinely used to assess response to treatment in some diseases, thereby enabling earlier therapeutic intervention in patients likely to relapse (Grimwade et al., 2009; Marin et al., 2011; Kuhn & Grimwade, 2012; Ong et al., 2012; Guilhot et al., 2012). While the clinical relevance of predictive biomarkers is perhaps clearer, these studies must similarly anticipate and plan for the context of therapeutic decision-making when designing evaluations of clinical utility (Simon, 2005a). Since clinical care is complex, standardizing the potential clinical pathways associated with various test results will reduce variation and enhance the ability of the study to assess the impact of test results on patient outcomes. The explicit description of the clinical pathways in terms of how the test results will be used is also important information for patients when considering whether to enroll in the study, as well as enabling the examination of the reasons for pathway selection.

To encourage transparent and complete reporting of study design and statistical analyses and promote reproducibility, reporting of test validation studies should utilize appropriate standards such as REMARK (Altman et al., 2012).

RECOMMENDATION 5: Outcome Measures for Clinical Utility

Studies to evaluate the clinical utility of MDx tests should include outcome measures that assess both potential benefits and harms of testing from the patient perspective, recognizing that these outcomes may occur at different time points and are the result of clinical management decisions guided by test results. Examples of typical outcome measures include clinical assessments of disease remission and progression, response to therapy, functional status as well as adverse events. Measures of benefits and harms should also routinely include patient-reported outcome measures, with the assurance that the selected measures are appropriate and validated for the clinical context. Clinical utility studies may also include important endpoints such as survival and downstream health care resource utilization; the decision to include these endpoints should be guided by the robustness of the existing evidence base regarding the specific clinical intervention prompted by the test result and its effects on relevant health outcomes. However process measures, such as changes in physician behavior, are typically insufficient to qualify as study endpoints, unless there exists a separate, robust body of credible evidence (as determined by widely accepted evidence review standards) linking clinical management decisions with relevant health outcomes.

Studies designed to report intended care plans following an MDx test are insufficient for demonstrating clinical utility.

Rationale: The primary clinical application for actionable MDx tests in oncology is to enhance risk stratification of patients for more precise risk classification, and targeting of interventions. The goal is to identify the subpopulation of patients that will obtain the optimal benefit/risk trade-off from an intervention so that ideally we can maximize the chances that any given patient will experience greater benefits than harms. These interventions may include watchful waiting, drug treatments, radiation, surgery or palliative care, with varying risk profiles depending on the clinical context and patient preferences. Characterization of the potential benefits and harms of testing should therefore be measured from the patient's perspective whenever possible, meaning the assessment of endpoints that reflect concepts that are perceived by, and of importance and relevance to, the patient. The specific term, "Patient-reported outcome" (PRO) has a more narrow definition, and a widely accepted version can be found in the FDA guidance for industry ("Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims") (FDA, 2007b) which defines a PRO as "any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else."

This position is entirely consistent with that of other major regulatory and policy-making organizations such as the National Cancer Institute (NCI, 2008), American Cancer Society (ACS, 2007), US Food and Drug Administration (Johnson et.al., 2003; FDA, 2007b), U.S. Centers for Medicare & Medicaid Services (CMS, 2007), and National Institutes of Health (NIH, 2007), that have emphasized the necessity of including PROs in addition to surrogate and clinical endpoints as part of cancer research. To that end, an entire EGD has been developed that describes specific standards for selecting non-regulatory PRO measures for clinical research (Basch et al., 2012). However to date, few studies have been conducted that fully describe the risks and benefits of MDx testing from the patient perspective, beyond traditional clinical measures of disease burden and progression. While results from MDx tests should be viewed in a similar fashion as other biomarkers (i.e., part of an overall clinical assessment contributing to informed patient decision-making) there are potential clinical benefits that have been described from MDx testing such as avoiding an ineffective therapy, switching more quickly to an effective therapy, or helping to choose among seemingly equal treatment options that may be sufficiently compelling if there is pre-existing evidence establishing the effectiveness of these treatments in comparable patient populations. However typically by themselves, these are intermediate outcome measures ultimately hypothesized to have an effect on key final outcome measures: patient health outcomes, as well as resource utilization. Therefore, a comprehensive assessment of the clinical utility of an MDx test requires an evaluation of the existing evidence base between associated health consequences

(benefits and harms) of clinical management decisions guided by the test results, prior to the selection of clinical utility study outcome measures. Typically evidence regarding patient perceptions of test result-guided management decisions (PROs) and health care system impacts (resource utilization) are not available in the diagnostics peer-reviewed literature and should be routinely incorporated as outcome measures in clinical utility studies.

RECOMMENDATION 6: Use of Randomized Controlled Trials for Clinical Utility

The clinical utility of an MDx biomarker should be assessed with randomized controlled trials that adequately evaluate the effectiveness of the clinical decision (treatment or other clinical pathway) relative to an appropriate control for both marker-positive and marker-negative patients. Enrichment designs which exclude from the study patients with a particular marker status should be avoided unless a clear and valid rationale exists for excluding non-marker status patients from study. Marker-based strategies which randomize patients to genomics-guided treatment vs. usual care partially duplicate the actions to be taken between the intervention and control arms, reduce statistical power and therefore are not optimal because they require larger sample sizes to demonstrate an effect of the MDx test.

Rationale: Definitive evaluation of the clinical utility of MDx tests used to guide patient management decisions will often require conducting randomized clinical trials (RCTs) (Simon, 2010). However, most RCTs of interventions focus on maximizing internal validity in order to accurately estimate the average treatment effect in a selected patient population. In the setting of MDx tests, much larger sample sizes are often needed to comprehensively assess the relationship between the biomarker and the treatment effect. Given the costs and time lag associated with conducting very large trials, alternative approaches that are applicable under certain circumstances are described elsewhere in this document. This recommendation focuses on appropriate RCT designs when RCTs must be used.

There are three major study design types that have been used to evaluate MDx biomarkers: biomarker-stratified designs, enrichment designs, and biomarker-strategy designs. While there are advantages and disadvantages to all of these RCT designs, in general designs that use the biomarker to guide the analysis are preferred over designs that use the biomarker to guide the treatment assignment (Friedlin et al., 2010).

The “all comers” marker-stratified design for evaluating MDx tests for clinical utility (Mandrekar & Sargent, 2009a; Mandrekar & Sargent, 2010; Maitournam & Simon, 2005; Simon, 2010; Simon, 2004; Simon, 2005b) is an approach in which all patients having the condition for which a management decision (therapy or other action) is required are enrolled into the study and tested to determine marker status. Each marker-defined group is then separately randomized to the treatments or decision pathways to be compared. This approach is preferred for evaluating MDx tests because it efficiently generates information on all relevant patients, including those who are marker negative. (Figure 5 provides a generalized example of this design for prognostic tests; Figure 6 illustrates this design for predictive tests.)

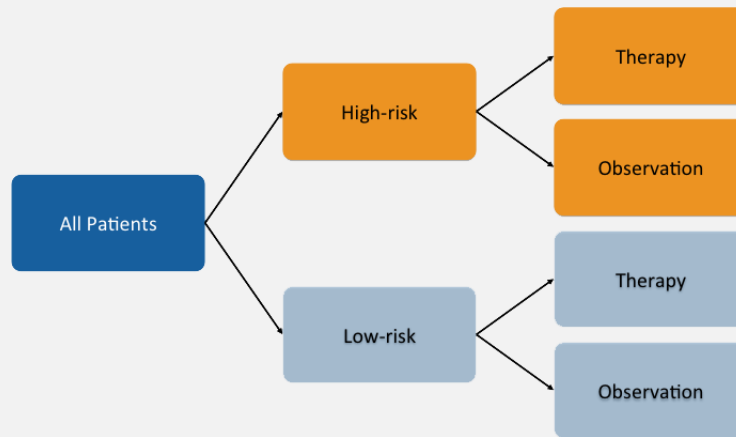


Figure 5: All Comers, Prospective Marker-stratified Design (Prognostic Test)

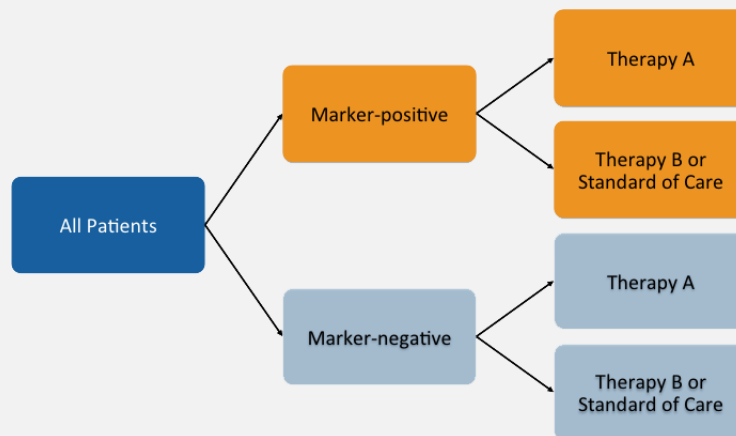


Figure 6: All Comers, Prospective Marker-stratified Design (Predictive Test)

The same design principle should be applied in situations where discovery of the molecular diagnostic biomarker has occurred after the development of an associated drug. In these cases, where adequate and appropriate tissue samples are available from a previous randomized controlled drug trial, a prospectively designed study using tissue samples from all comers is a useful approach to consider (See Recommendation 7 and Figure 9 for more detail).

In some cases, e.g., where compelling evidence exists that marker-negative patients cannot benefit from a treatment, or when a group of responders has been identified for further study within an otherwise highly heterogeneous population, enrichment designs are useful to focus on a specific group of interest (FDA, 2012a). In this approach for marker-related drug development, patients are initially tested and only marker-positive (or negative) subjects are included in the study (Figure 7) (Mandrekar & Sargent, 2009a; Mandrekar & Sargent, 2009b; Mandrekar & Sargent, 2010). While informative for targeting specific populations, enrichment designs are sometimes seen as an attractive option because they can require fewer randomized patients than non-targeted studies (Simon, 2004; Maitournam & Simon, 2005). Generally speaking, however, the approach is only justified in cases where the biologic rationale and preliminary evidence that only one group benefits is compelling enough that equipoise does not pertain between the existing alternatives for all patients, making it unethical to randomize treatment options to all marker-based groups. Otherwise, enrichment studies are not recommended because no information is obtained on the group excluded from the study. *Hence, enrichment designs cannot definitively establish the predictive ability of a marker.*

The third category of clinical trial design, the biomarker strategy design, is to study genomics-guided treatment vs. usual care in which the patients randomized to usual care are not tested. This design is sometimes described as the “gold standard” because it attempts to replicate what would occur in clinical practice. However, in this strategy some patients receiving MDx-guided therapy receive the same treatment (standard of care) as patients in the standard therapy arm, which dilutes the ability to see a treatment effect (see Figure 8) (Micheel et al., 2012; IOM, 2011). While this design is very flexible, the same objectives can typically be achieved with fewer patients using the marker-stratified design described above. Given the larger sample size required to demonstrate a difference between study arms, the biomarker strategy design is not preferred.

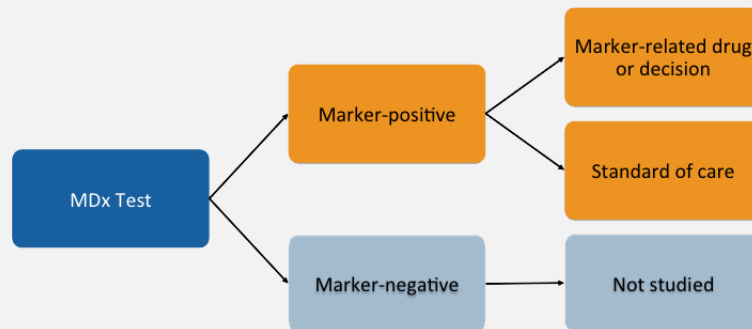


Figure 7: Marker Enrichment Design

Not recommended except under specific circumstances; no information is obtained on excluded group.

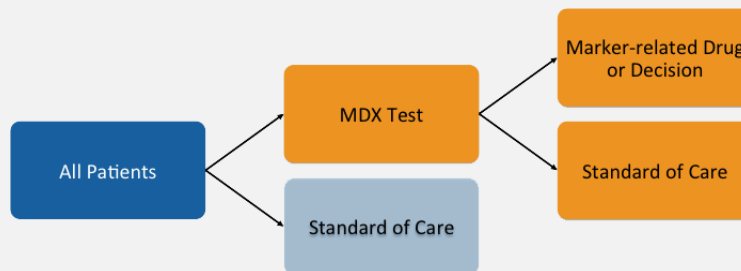


Figure 8: Biomarker Strategy Design

Not preferred because the approach reduces statistical power, given that patients in both study arms receive the “standard of care” as their intervention.

RECOMMENDATION 7: Using Prospective-retrospective Analysis of Previously Conducted RCTs

If an appropriately designed, powered and conducted clinical trial with banked biospecimens exists, then a properly conducted prospective-retrospective study is adequate evidence of clinical utility. Replication of study results (second study) and

pooling of biospecimen samples from comparable RCT's are two approaches to address limitations related to causal inference and insufficient sample sizes.

Rationale: Typically the best setting to evaluate the clinical utility of an actionable MDx test in oncology is an RCT and the most efficient design is an RCT of standard therapy vs. a targeted where biomarker status is obtained on all patients but is not used to guide treatment (see Recommendation #6). While ideally biomarker development would occur in parallel with drug development, commonly biomarker development lags behind for many different scientific and technical reasons (Taube et al., 2009). Therefore it is frequently the case that retrospective analyses are conducted using archived tissues several months to years after a cancer therapy has been approved. Although these studies are often published in the peer-reviewed literature, many experts have judged published retrospective biomarker studies as demonstrating insufficient evidence to establish the clinical utility of the markers studied (Simon et al., 2009; Henry & Hayes, 2006; Ransohoff, 2004). Recognizing these limitations, methodological recommendations for appropriately designing studies to demonstrate clinical utility using previously collected biospecimens have been described. (Simon, Paik, and Hayes, 2009; McShane, 2012; Polley, et al., 2013).

For example, where a new marker has been identified after definitive trials have demonstrated patient benefit for a specific agent (or class of agent or type of treatment), test developers should determine the availability of high-quality archived tissue specimens from these previously completed RCTs. To ensure the appropriate use of a “prospective-retrospective” study design to evaluate the clinical utility of a new biomarker, several conditions must be present to ensure that this approach is of sufficient scientific rigor to convincingly demonstrate clinical utility. Simon, Paik, and Hayes (2009) specify that: 1) enough archived tissue must be available for an appropriately powered study and to assure that patients included in the biomarker evaluation are representative of the patients in the source RCT (as a suggested rule-of-thumb, tissue from at least two-thirds of patients in the RCT should be available for inclusion in the prospective-retrospective analysis); 2) the analytic validity of the test must be well established to assure that results from archived tissues resemble the results that would have been obtained from real-time tissue collection; 3) the analysis plan for the biomarker study must be completely pre-specified; 4) the results must be validated in at least one or more similarly designed studies using the same assay techniques.

Our project team debated the pros and cons of requiring that a prospective-retrospective study be replicated in order to demonstrate clinical utility. While replication would always be viewed positively, the TWG felt that if a single, properly designed and adequately powered, prospective-retrospective study had positive results, this would be considered adequate evidence of clinical utility. More recently, others have suggested

the importance of ensuring adequate sample size to test for both quantitative and qualitative interactions (McShane, 2012). The presence of a treatment by marker interaction means that the treatment effect (difference in clinical outcome between study arms) differs depending on the patient’s marker status; a treatment by marker interaction is most clinically relevant when it is a qualitative interaction. Since the sample size required to test for interactions typically exceeds that required to test for a therapeutic effect, it may be necessary to pool samples for across studies. Successful pooling assumes that the clinical settings are similar across the different trials and that the specimens were all tested with the same assay or at least one that is comparable. (Polley, et al., 2013)

A clinically important example of prospective-retrospective of evaluation can be seen in KRAS testing for anti-epidermal growth factor receptor (EGFR) antibody therapy for metastatic colorectal cancer patients (Patterson, et al., 2011; Burns, et al., 2010). The efficacy of the anti-EGFR agents cetuximab and panitumumab in metastatic colorectal cancer was established in separate randomized controlled trials (Jonker et al., 2007; Van Cutsem et al., 2007). Then the association of KRAS mutation status with treatment response was subsequently established through non-concurrent subgroup analyses using archived tissue samples from these efficacy trials (Amado et al., 2008; Karapetis et al., 2008). KRAS testing prior to initiation of anti-EGFR therapy is now considered the standard of care.

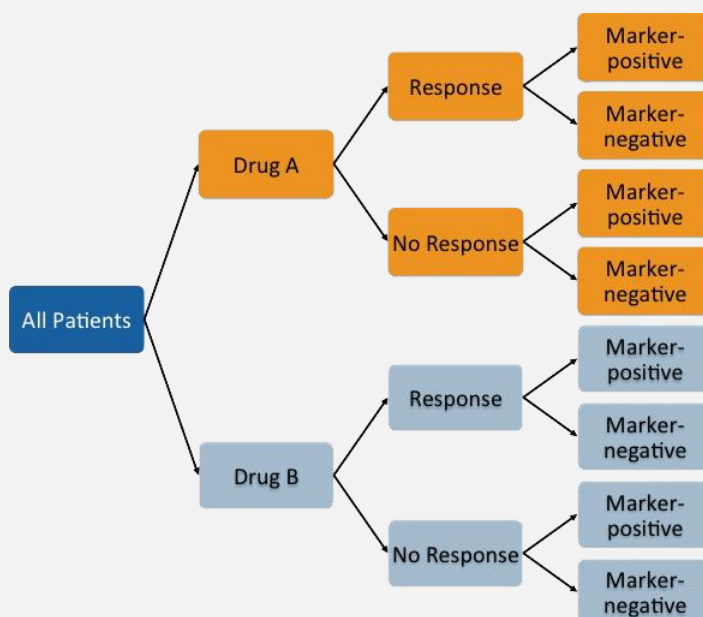


Figure 9: Prospective-retrospective RCT Design

Drug is tested in RCT first and marker-status is determined retrospectively from tissue samples. Recommended for situations in which marker was not known when drug was first developed. Can also be used for independent validation.

RECOMMENDATION 8: Using Single-arm Studies RCTs

Single-arm studies can be used to establish the clinical utility of an MDx test in cases where all of the following conditions are met: 1) the MDx test is being developed to be used with a drug that has already been FDA-approved on the basis of pivotal trials of a general population with regulatory endpoints such as survival or progression-free survival; 2) adequate archived tissue samples are not available to conduct a prospective-retrospective trial to assess clinical utility of the MDx test; 3) it is feasible to use response, variably defined as complete or overall response, as an endpoint in the single-arm study; and 4) there exists comparable response data in a non-contemporaneous comparative cohort.

Rationale: MDx tests identifying a subset of patients who benefit differentially from a drug treatment are sometimes developed *after* the drug in question has been FDA-approved for marketing. In these cases, the drug in question has been tested in randomized controlled trials (RCTs) in a broad patient population (i.e., meeting eligibility requirements related to stage of disease, demographic variables, and other common criteria, but not screened according to the result of an MDx test) using a regulatory endpoint such as survival or progression-free survival (PFS). Since the drug is approved for use across the whole population of patients meeting the disease-specific indication given in the labeling, it would not be ethical or practicable to conduct subsequent RCTs in which a control group is denied the approved therapy. If adequate tissue samples have been preserved from previously done registration trials, it would be desirable to design a prospective-retrospective trial of the MDx test using preserved tissue. If insufficient archived tissue is available, however, new RCTs would not be a viable option for clinically testing a new MDx test purporting to identify some subset of patients who benefit differentially from the treatment.

An alternative in such cases is to conduct a single-arm study. Without a control arm, survival or progression-free survival would not be interpretable endpoints. However, tumor shrinkage is typically taken to be a sign that a patient will benefit from a therapy in oncology. In addition, although there is not a contemporaneous control arm for comparison, in many cases the study can be interpreted in the context of the response of a non-contemporaneous cohort. Under these circumstances, tumor response could be

used as an endpoint to evaluate differential benefit between test-positive and test-negative patients in a single-arm study to establish the clinical utility of the test. Single-arm studies of this type are not as robust as RCTs, since they only provide information on the test-positive patients, not the test negative patients. In some cases, such as erlotinib, the response rate is much greater in marker-positive patients, but marker-negative patients still have a meaningful benefit, which can only be demonstrated by randomized clinical trials. For this reason, **the test-negative patients cannot be assumed not to benefit from the treatment.** Nevertheless, the marker-based differential tumor response can provide useful data to the clinician to be used in the context of other relevant information to create an individual treatment plan.

RECOMMENDATION 9: Using Longitudinal Observational Studies

Under limited, specified circumstances, *longitudinal* observational study designs such as prospective cohort studies, patient registries that explicitly include comparators, and multiple group, pretest/posttest designs (also called quasi-experimental, difference-in-difference design; regression discontinuity design) are acceptable options for assessing the clinical utility of MDx tests, provided that a compelling rationale for not doing an RCT is addressed (examples below), efforts to minimize confounding are documented, and good research practices for prospective observational studies are followed, including public registration of studies. Since the necessary parameters for evaluating the clinical utility of MDx tests (e.g., clinical characteristics of patients, test findings and interpretation, subsequent care and patient outcomes) are typically not found in secondary databases (including most electronic health records), the pursuit of retrospective observational studies is generally not adequate. Prospective observational studies may include the use of secondary databases as one component of the data collection effort, but must also include prospective data collection efforts to obtain the missing data or develop validated approaches to approximate these data elements from the existing secondary data.

Rationale: Given the time, cost and limitations with respect to generalizability of traditional RCTs, there has been a longstanding interest in developing a broader range of methods for evaluating the effects of MDx testing on patient outcomes. The decision to pursue an observational study rather than an RCT is driven by careful consideration of state of clinical equipoise for the MDx test of interest (i.e., there is genuine uncertainty on the part of the expert medical community regarding the preferred clinical pathway)

and whether the proposed study design and analysis plan will sufficiently address potential problems with confounding and bias (Berger et al., 2012). The key design characteristic of a prospective observational study is one where participants are not randomized or otherwise preassigned to the MDx test decision, but are followed longitudinally for both subsequent clinical interventions and patient outcomes. This design is in contrast to a retrospective observational study, which relies on existing data sets and both exposure (MDx test) and clinical outcomes have occurred in the past. Due to reliance on existing databases, retrospective observational studies typically have advantages in terms of cost and speed of execution; however because of an absence of critical data elements for evaluation of the clinical utility of MDx tests, this approach is discouraged at this time. Efforts to improve the specificity of test codes (AMA, 2013) and include test results in linked electronic health records that also include patient reported outcomes may improve the suitability of secondary databases in the future. We support ongoing multi-stakeholder efforts such as those led by the American Society for Clinical Oncology to implement oncology and MDx electronic data standards (e.g. Hurley et al., 2010).

Nevertheless, significant concerns remain regarding criteria for selecting a prospective observational study design over an RCT for the purpose of demonstrating clinical utility, given that the emphasis on “real-world” interventions and outcomes generally increases the potential for time-varying and time invariant confounding, as well as bias. There are a number of best practices that should be adopted to minimize these threats to validity, including assuming from the outset that an observational study approximates a randomized study in its overall objective of causal inference and therefore requires a full protocol with corresponding hypotheses and specified intervention groups, definitions of outcome measures as well as subgroups, power calculations and an analysis plan that describes how potential confounding, missing data and loss to follow-up will be handled. (Rubin, 2007). A complete user’s guide on best practices for designing observational studies has been prepared by AHRQ (Velentgas et. al, 2013). In addition, a thorough description of good research practices for prospective observational studies has been published by a task force of experts (Berger et al., 2012) as well as a user’s guide for establishing registries to evaluate patient outcomes (Gliklich & Dreyer, 2007). Of particular relevance for studies of MDx tests is the issue of heterogeneity of treatment effects (HTE), defined as the nonrandom variability in treatment effects due to patient factors, in this case due to genomic or other molecular variation related to the patient’s tumor. A checklist of key considerations regarding how to plan the HTE/subgroup analysis section of an observational study has been developed in one chapter in the aforementioned AHRQ methods guide. (Varadhan & Seeger, 2013).

By way of comparing our two recommended approaches, the quasi-experimental designs (pretest-posttest) collect outcome data before and after the intervention (MDx test) for at least two groups, using individuals as their own controls, and are therefore robust to

time-invariant unmeasured confounding. The prospective cohort design facilitates broad enrollment criteria and therefore enables examination of treatment heterogeneity, but like the former design is also vulnerable to time-varying unmeasured confounding. For example, disease severity is a time-varying confounder — it evolves over time and is likely to be linked to the indication for testing, whereas a patient’s socioeconomic status is time-invariant (at least for the study duration) and also likely to be linked to the decision to undergo testing. Assuming that good research practices are observed, factors that influence the decision to pursue an observational study include: hypothesized large effect size ($RR > 2$); a treatment strategy likely to evolve over time because treatment depends on clinical course and disease activity and therefore clinicians are reluctant to randomize; investigators are interested in studying how MDx test results are implemented in real world settings, assuming there are a plausible range of downstream clinical choices based on test results; investigators are interested in studying patients’ willingness/preferences to act on test results; there is a need to study long-term outcomes such as survival; very large sample sizes are required to assess clinical utility. Efforts to minimize bias should extend to practices related to study execution, interpretation and reporting (Berger et al., 2012, GRACE, 2010; Velentgas et. al, 2013). Also, there have been calls in the US to register prospective CER studies as well as patient registries (Gliklich & Dreyer, 2007), but now there has been a specific recommendation to register all tumor marker studies as well (McShane & Hayes, 2012), which we support and which should help ensure the quality of studies over time.

RECOMMENDATION 10: Using Decision-Analytic Modeling Techniques

Based on initial scenario modeling, formal decision-analytic modeling techniques can be used to elucidate the relationship between test results, corresponding clinical pathways and downstream patient outcomes in cases where an MDx test has established evidence of clinical validity and plausible evidence of clinical utility based on initial scenario modeling (a simplified approach to decision analysis that typically includes outcomes evaluated under 3 scenarios: base case, best case, worst case). Decision-analytic models are useful in the common situation where there is no direct evidence of clinical utility, as they provide explicit estimates of the likely effects of clinically validated test results on patient outcomes by linking separate sources of evidence, including quantifying the relationship between surrogate outcome measures and final patient outcomes. Models should include all patient-relevant benefits and harms related to the duration and quality of remaining life. Summary measures such as clinical events, life expectancy and quality-adjusted life years represent appropriate modeling outcome measures. Good modeling practices for diagnostic tests have been

published and should be followed; these methods are labor and time intensive and are not recommended when there is a high degree of uncertainty about the underlying disease process, lack of a clinical intervention with known benefits, or when there is high uncertainty about the link between test results and the effectiveness of interventions.

Rationale: Models, broadly defined as “mathematical frameworks that facilitate estimation of the consequences of healthcare decisions” (Caro et al., 2012) are particularly prevalent in the evaluation of the diagnostics literature precisely because there is a paucity of published direct evidence of their clinical utility. Decision-analytic modeling is a descriptive term that is used in this document to describe a type of model that is used to depict a common clinical scenario in MDx testing, however other model types such as state-transition models or discrete event simulations may be appropriate depending on the clinical situation (Roberts et al., 2012). Our goal is not to provide guidance about how to develop valid models, as best practices have been described elsewhere (Caro et al., 2012; Trikalinos et al., 2009; Samson & Schoelles, 2012), but rather to recommend circumstances when modeling would be an appropriate method for assessing the clinical utility of an MDx test.

Given that model development can be resource intensive, researchers typically first assess the desirability of conducting a formal decision modeling exercise by first developing a simple decision model called a scenario model that consists of a simplified decision tree and a series of “what if” scenarios intended to provide a quantitative assessment of the general likelihood that an MDx test will demonstrate clinical utility (Veenstra et al., 2013). If a flow diagram (see above) has already been developed, then the key parameters and assumptions under three scenarios (base case, best case and worst case) should be revisited with key stakeholders (patients, clinicians and payers) and the outcomes estimated for each case. The purpose of this step is to identify MDx tests that fail to meet a stakeholder-driven threshold for plausible evidence of clinical utility. Researchers with extensive experience with MDx test modeling describe common limitations for further formal decision modeling, such as including insufficient or unreliable data on the prevalence of genomic variants; insufficient data on the effectiveness of the test-driven interventions; and nontransferability of test performance estimates across populations and settings (Trikalinos et al., 2009; Veenstra et al., 2013). For MDx tests that cross the plausibility threshold, modeling techniques are used to project the overall downstream health outcomes that in most instances may not be available even within the context of RCTs, due to limited follow-up, highly selected patient populations, and/or small sample sizes.

A comprehensive assessment of the clinical utility of an MDx test should include all patient-relevant benefits and harms related to the duration and quality of the remaining

life (Trikalinos et al., 2009) such as modeled estimates of clinical events, life expectancy and quality-adjusted life-years, as appropriate (Veenstra et al., 2013). Modeling in this context also refers to establishing the link between surrogate outcome measures and final outcome measures, assuming that a separate body of evidence exists linking the two. For example, if clinical trials have already established the link between PROs and survival or use of an approved anti-cancer treatment and patient outcomes, then modeling can provide quantitative estimates of the overall benefits and harms for a given population of patients. Alternatively, there may be data from separate studies demonstrating the relationship between biomarker status, various steps in the care pathway and patient outcomes – these may be quantitatively linked through modeling to provide estimates of the net benefit to patients (Elkin et al., 2004). In both of these situations, decision-analytic models provide a useful framework for evaluating the benefits and harms of MDx tests by constructing the link between clinical validity measures and patient outcomes using data from different sources, while simultaneously making explicit the key parameters influencing outcomes and demonstrating the impact of parameter uncertainty through the use of sensitivity analyses (Yang, 2012).

One cancer-specific example is the EGAPP assessment of the clinical utility of mismatch repair (MMR) mutation testing in Lynch syndrome, a hereditary predisposition to colorectal cancer (CRC) and some other malignancies (EGAPP, 2009). Lynch syndrome results from the autosomal dominant inheritance of a germline MMR gene mutation. The four major genes of interest are *MLH1*, *MSH2*, *MSH6*, and *PMS2*. The risk of CRC in subjects with Lynch syndrome is high for both a second primary CRC in a diagnosed patient (approximately 16% within 10 years) and a new cancer in a first- or second-degree relative with Lynch syndrome (approximately 45% for men and 35% for women by age 70). The question the EGAPP review sought to answer was: “Does risk assessment and MMR gene mutation testing in individuals with newly diagnosed CRC lead to improved outcomes for the patient or relatives, or is it useful in medical, personal, or public health decision-making?”² The review group found adequate evidence of analytic validity and clinical validity, but not studies which directly answered the clinical utility question posed. Nevertheless, the group was able to construct a chain of evidence (Lewin Group, 2009) from studies that individually examined the following components of clinical utility assessment: testing uptake rates, adherence to recommended surveillance activities, number of relatives approachable, harms associated with additional follow-up, and effectiveness of routine colonoscopy. Although there was insufficient evidence to recommend a specific testing strategy (out of a choice of four strategies), assessment of these individual components created a chain of indirect but convincing evidence that offering genetic testing to relatives of newly diagnosed patients with Lynch syndrome would reduce morbidity and mortality in relatives due to the resulting effective preventive clinical management for the relatives. See reference for a general model and

² EGAPP, 2000 p. 36.

additional case examples.

IMPLEMENTATION BARRIERS

The following section discusses key challenges to implementing the methodological recommendations discussed in previous sections of this EGD, suggesting a range of possible solutions to overcoming these hurdles. In reviewing these discussion points, some TWG members expressed the opinion that these non-methodological recommendations were qualitatively different and potentially outside of the scope of a document designed to primarily guide researchers in their study design efforts. Other TWG members believed it was important to acknowledge the larger environment in which test development takes place and address the current barriers to generation of evidence of clinical validity and clinical utility. All participants recognized that while the target audience for the EGD is primarily researchers and test developers, the target audience for these broader solutions also include policymakers. For the CMTP project team, it seemed clear that the current evidence gaps exist not merely because of a need for clearer methods guidance, but also because the regulatory and commercial milieu in which tests are developed tends to create disincentives to generating the needed data. Therefore, CMTP opted to include this section to underscore the need for both methodological guidance and broader problem-solving that reflects multi-stakeholder partnerships. The following position statements should not be construed as applicable in every situation; to the contrary, evidence development for each new test will pose specific challenges requiring carefully tailored solutions which may employ one or more of the approaches discussed here, or may call for other innovative approaches.

POSITION STATEMENT 1: Encouraging Public-Private Collaborations

We encourage sustainable new public-private collaborations in order to ensure the efficient implementation of clinical utility studies for specific MDx tests. The goal is to broaden the stakeholder engagement model that served as the foundation for the development of MDx evidence standards to larger groups of decision-makers. This facilitates a shared understanding of evidence thresholds as they apply to specific MDx tests, and ensures that individual clinical utility studies are maximally informative for clinical and coverage decision-making. In particular, we suggest that entities such as diagnostic and pharmaceutical companies, payers, cancer centers, accountable care organizations (ACOs), universities and professional societies collaborate to develop an interoperable research infrastructure and apply the EGD recommendations to individual study proposals and protocols. Over time, this will lower the barriers to developing evidence of clinical utility by providing consistent, predictable and uniform evidence standards for researchers and test developers, provide real-world

opportunities for feedback and refinement of the standards, while providing sufficient flexibility for decision-makers to tailor their application to specific MDx tests.

Rationale: Currently there is a misalignment between the study endpoints selected by researchers for many clinical studies involving molecular diagnostic tests and the outcomes that are relevant and acceptable to decision-makers such as payers. For industry-based researchers, this misalignment typically occurs due to a lack of regulatory requirements for demonstrating evidence of clinical utility, the costs associated with conducting clinical trials powered to demonstrate clinical utility and, most importantly, the lack of certainty regarding evidentiary requirements for coverage and reimbursement of new MDx tests. For academic researchers, federal agencies traditionally have not funded CU studies and the principal investigator-driven model of conducting research has not trained and rewarded investigators for designing studies to meet the information needs of decision-makers. Today, a few examples of public investment in assessing the clinical utility of MDx tests exist, such as large RCTs to evaluate Oncotype Dx in node negative (the TailorRX trial) (Zujewski & Kamin, 2008) and node positive (the RxPONDER trial) (Ramsey et. al., 2013) breast cancer patients as well as proof-of-concept stakeholder-driven models of conducting comparative effectiveness research for high priority MDx tests (the CANCERGEN project) (Thariani et al., 2012). The lessons learned from these early examples is that real-world coverage decisions often are out-of-sync with research timelines and that multi-stakeholder involvement is essential for both assessing study design considerations and for implementing these studies on a wider scale going forward. While greater clarity regarding evidentiary standards for clinical utility and coverage decision-making is a critical first step, there needs to be ongoing collaboration between public and private entities such as diagnostic and pharmaceutical companies, payers, cancer centers, ACOs, universities and professional societies to assure development of a research infrastructure able to accommodate in an economically sustainable manner a set of uniform standards and to effectively implement these standards in the design of specific clinical utility studies for particular MDx tests. These public-private collaborations would involve a range of intellectual, infrastructure and financial contributions to ensure that the CU standards as described in this EGD are implemented pragmatically and effectively, in a fashion such that individual studies are likely to meet the information needs of payers, clinicians and patients. These collaborations would focus on sharing knowledge and reducing the risk that studies undertaken by private or public sponsors would not be informative for policy decision-making. Whenever possible, these public-private collaborations would take advantage of infrastructure and tools afforded by the cancer cooperative groups and ongoing investments in registries and electronic medical records (particularly ones that facilitate the use of electronic medical records for clinical research) and other methods of streamlining clinical research.

POSITION STATEMENT 2: Novel Reimbursement Policy Approaches to Promote Clinical Utility Studies

We support the development and use of novel reimbursement policy approaches to promote clinical utility evidence generation for molecular diagnostic tests and other medical devices and drugs. Managed entry schemes encompass a broad range of policy tools that provide the flexibility to payers to cover innovative, emerging molecular diagnostic tests while generating valid evidence on the relative benefits and risks of these tests while they are used in clinical practice. Among the possible tools to be considered on a case-by-case basis are FDA-CMS parallel review and adaptive licensing (for companion diagnostics and in vitro diagnostic tests undergoing FDA review) and performance-based risk-sharing arrangements (potentially applicable to both LDTs and in vitro diagnostic tests undergoing FDA review), including the provision of coverage for patients in well-designed clinical trials to gather CU evidence for clinically promising MDx tests (coverage with evidence development).

Rationale: Currently a two-track system exists for market entry of MDx tests. Test products intended to be marketed as test kits undergo FDA review as in vitro diagnostic devices as do tests developed as decision-making tools to guide the clinical use of a specific drug (companion diagnostics). By contrast, laboratory-developed tests (LDTs) are created and used solely by the developers of the test.

Currently, most MDx tests are developed as LDTs which are not required to undergo FDA device review (Javitt & Hudson, 2006; SACGHS, 2008a; SACGHS, 2008b). Hence, the developers of LDTs have little incentive to develop evidence of clinical utility, since it is not required for FDA clearance and the trials can be costly. The FDA has claimed authority to regulate LDTs, but thus far has not moved to require review (FDA, 2010). One consequence of this mode of development is that while the analytic validity is evaluated to satisfy requirements under the Clinical Laboratories Improvement Act (CLIA), and some information may be generated on the association between the marker and the physiological state of interest as part of the development process (clinical validity), any demonstration of the impact of the test on clinical decision-making or patient outcomes (clinical utility) is never formally generated, or only comes through slow accretion of clinical use (Woodcock, 2010).

Nevertheless, as these tests are gradually taken up into clinical practice, guideline developers and practitioners need information on how best to use the tests clinically, and US healthcare payers wish to see evidence of a clinical benefit to patients before agreeing to provide reimbursement for routine clinical use (Schulman & Tunis, 2010). Availability of more complete evidence of utility tends to promote more rapid clinical uptake of new tests, therefore benefitting patients more quickly. Hence, alignment of

incentives is needed to promote the conduct of studies that will inform payer, patient and clinician decision-making and will minimize the potential for inappropriate use, risks and economic harms while allowing limited access to possibly beneficial new tests.

It may be possible to derive the needed information through evidence-based reviews and development of indirect evidence chains (see Recommendation 9). However, if the required data are not available, other tools to bridge this data gap are still required. FDA-centered policy approaches such as FDA-CMS parallel review (FDA, 2011b; CMS, 2010; Messner and Tunis 2012)(a pilot program underway to coordinate the clearance process of FDA with the national coverage determination process of the Centers for Medicare and Medicaid) or adaptive licensing (Eichler, et al. 2012) (gradually staged, evidence-oriented, performance-based market entry, currently being attempted in some non-US jurisdictions) would not be applicable, since either approach requires FDA review to take place.

Another possible approach to foster the generation of clinical utility evidence is to use one of a range of performance-based risk-sharing arrangements (PBRsAs) (Carlson et.al. 2010). PBRsAs entail contractual arrangements between payers and the diagnostic device developer such that the price, level or nature of reimbursement for a particular technology is tied to future measures of clinical endpoints and real world effectiveness of technologies (ISPOR, 2011). Within this constellation of arrangements, one possibility in some situations is for the payer to provide coverage for a limited number of patients who agree to enroll in a clinical study to provide the additional evidence required (an approach called coverage with evidence development). Among other criteria which may apply on a case-specific basis: (i) Decision-makers would need to judge the existing medical literature insufficient to answer key questions, (ii) the MDx test must be seen as having the potential to provide substantial improvement in public health outcomes, (iii) there must be adequate evidence of CV and AV, and a clear biologic rationale for CU, (iv) data collection must be the best solution to resolve the uncertainty regarding the safety and effectiveness of the MDx test, (v) the primary concern should be uncertainty regarding clinical utility, (vi) stakeholders should agree that the evidence development can be achieved in a timely manner, and (vii) market factors suggest that in the absence of a PBRSA initiative, relevant evidence would not become available until after the technology is already in wide use. Additionally, should the MDx test prove not to hold the promise of clinical utility anticipated, there should be a clear pathway for the payer to withdraw reimbursement support from the test.

POSITION STATEMENT 3: Enhancing MDx Test Comprehension for Healthcare Professionals

We support initiatives that enable healthcare professionals to accurately interpret and communicate the results of molecular diagnostic testing to patients and their caregivers. Strategies include providing Continuing Medical Education credits (CME) for MDx-related training, as well as engaging professional societies to develop practice guidelines specifying the use and interpretation of MDx test results. In addition, test developers must work with both clinicians and patient advocates to design reporting templates that can be informative to all stakeholders, including patients. Therefore, we recommend that these groups collaborate to develop test reports that are maximally useful to patients, caregivers, and health care professionals.

Rationale: The value of an MDx test is only realized when used appropriately in practice; however the rapid pace of scientific advances and the complexity of molecular testing complicate the interpretation of MDx tests (Green & Guyer, 2011). Physicians report a knowledge deficit with respect to the use of MDx tests in practice, although specialists such as oncologists typically report higher levels of comfort with interpreting these tests than primary care physicians (United Healthcare, 2012). Although test developers have recognized this need and are developing reporting templates to improve interpretability of test results for clinicians (Scheuner et al., 2012; Genoptix, 2012; CoreGenomics, 2012), there is a considerable gap in communicating results of MDx testing in a format that can be interpreted by patients. Patients need access to information about MDx testing that is accurate, reliable, and understandable, and that clearly describes the implications of test results for their own prognosis and treatment options (Engstrom et al., 2011). Patient understanding of treatment options and the risk and benefits associated with each option is the cornerstone of shared decision-making and a necessary component of patient-centered care (Coulter, 2012). Therefore, efforts that enable patients and physicians to use test results effectively and appropriately to make treatment choices is essential for rapid adoption of an MDx test in clinical practice. In addition, the utility of an MDx test continues to evolve as new information becomes available. Therefore, mechanisms to ensure that the most up-to-date information is available to patients and healthcare professionals will enable appropriate use of the test in clinical practice.

REFERENCES

- Agency for Healthcare Research and Quality (AHRQ). AHRQ Effective Health Care Program: Glossary of Terms. 2013. Retrieved at <<http://www.effectivehealthcare.ahrq.gov/index.cfm/glossary-of-terms>>.
- American Cancer Society (ACS). 2015 Goals of the American Cancer Society. 2007. Retrieved at: http://www.cancer.org/docroot/COM/content/div_PA/COM_4_2X_Volunteer_11087.asp
- American Medical Association (AMA). CPT 2013 Professional Edition. Item#: EP888813 ed: American Medical Association 2013. " Accessed 22 April 2013. Retrieved at: < https://commerce.ama-assn.org/store/catalog/productDetail.jsp?product_id=prod1180004&navAction=push>.
- Altman DG, McShane LM, Sauerbrei W, et al. Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): Explanation and Elaboration. *BMC Medicine* 2012; 10:51.
- Altman DG, Vergouwe Y, Royston P, et al. Prognosis and prognostic research: Validating a prognostic model. *BMJ* 2009; 338:b605.
- Amado RG, Wolf M, Peeters M, et al. Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. *J Clin Oncol* 2008; 26:1626-34.
- American Cancer Society (ACS). "2015 Goals of the American Cancer Society." 2007. Accessed 17 November 2012. Retrieved at: <http://www.cancer.org/docroot/COM/content/div_PA/COM_4_2X_Volunteer_11087.asp>.
- Atkinson AJ, Colburn WA, DeGruttola VG, et al. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clin Pharmacol Ther* 2001; 69:89-95.
- Basch E, Abernathy A, Mullins D, et al. Recommendations for Incorporating Patient-Reported Outcomes into the Design of Clinical Trials in Adult Oncology: Effectiveness Guidance Document. 2012. Center for Medical Technology Policy. Baltimore, Maryland. Retrieved at: <<http://www.cmtpnnet.org/effectiveness-guidance-documents/oncology-pro/>>.
- Berger ML, Dreyer N, Anderson F, et al. Prospective observational studies to assess comparative effectiveness: The ISPOR good research practices task force report. *Value in Health* 2012; 15:217-230.
- Betsou F, Lehmann S, Ashton G, et al. Standard preanalytical coding for biospecimens: Defining the sample Preanalytical code. *Cancer Epidemiol Biomarkers Prev* 2010; 19:1004-1011.
- Blue Cross Blue Shield Association Technology Evaluation Center (BCBSA TEC). "Special Report: Companion Diagnostics—Example of BRAF Gene Mutation Testing to Select Patients with Melanoma for Treatment with BRAF Kinase Inhibitors." 2011. Retrieved at: <http://www.bcbs.com/blueresources/tec/vols/26/26_07.pdf>.
- Bossuyt PM, Reitsma JB, Bruns DE, et al. The STARD Statement for Reporting Studies of Diagnostic Accuracy: Explanation and Elaboration. *Ann Intern Med* 2003; 138:W1-W12.
- Burns DK, Hughes AR, Power A, et al. Designing pharmacogenomics studies to be fit for purpose. *Pharmacogenomics* 2010;11(12):1657-1667.
- Campollo Rivas O. Mathematical modelling in medicine and biology: Theoretical basis and fundamentals. *Rev Invest Clin* 1994; 46(4):307-21.
- Carlos R, Gareen I, Gatsonis C, et al. Standards in the design, conduct and evaluation of diagnostic testing for use in patient centered outcomes research. 15 March 2012. Washington, D.C.
- Carlson JJ, Sullivan SD, Garrison LP, et al. Linking payment to health outcomes: A taxonomy and examination of performance-based reimbursement schemes between healthcare payers and manufacturers. *Health Policy* 2010; 96:179-90.

- Caro JJ, Briggs AH, Siebert U, et al. Modeling good research practices—overview: A report of the ISPOR_SMDM modeling good research practices task force-1. *Value in Health* 2012; 15:796-803.
- Center for Medical Technology Policy (CMTP). Molecular Diagnostics Technical Working Group Conference Call. January 23, 2012.
- Center for Medical Technology Policy (CMTP). CMTP Molecular Diagnostics in Oncology Technical Working Group Meeting. September 2011a. Baltimore, MD.
- Center for Medical Technology Policy (CMTP). Patricia Deverka and Donna Messner interviewers. Personal interview with Robert Epstein. April 27, 2011b.
- Center for Medical Technology Policy (CMTP). Patricia Deverka and Donna Messner interviewers. Personal interview with Richard Simon. August 16, 2011c.
- Center for Medical Technology Policy (CMTP). Technical Working Group (TWG) Meeting: Molecular Diagnostics in Oncology. 2011d. Baltimore, MD.
- Centers for Disease Control and Prevention (CDC). Good Laboratory Practices for Molecular Genetic Testing for Heritable Diseases and Conditions: Morbidity and Mortality Weekly Report. Recommendations and Reports. 2009. Retrieved at: <<http://www.cdc.gov/mmwr/pdf/rr/rr5806.pdf>>.
- Centers for Medicare & Medicaid Services (CMS). Medicare Launches Efforts to Improve Care for Cancer Patients: New Coverage, Better Evidence, and New Support for Improving Quality of Care. 2007. Accessed at: <http://www.cms.hhs.gov/media/press/release.asp?Counter_1244>.
- Centers for Medicare and Medicaid Services and Food and Drug Administration (CMS-FDA). Parallel review of medical products. In: Department of Health and Human Services. 2010.
- Chalkidou K, Tunis S, Whicher D, et al. The role of pragmatic randomized controlled trials (pRCTs) in comparative effectiveness research. *Clin Trials* 2012; 9:436.
- Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 2007; 115:928-35.
- CoreGenomics: Foundation Medicine's Cancer Genomics Tests. 2012. Retrieved at: <<http://core-genomics.blogspot.com/>>.
- Coulter A. Patient engagement – What works? *J Ambulatory Care Manage* 2012; 35(2):80-89.
- Deverka P, Doksum T, Carlson R. Integrating molecular medicine into the U.S. health-care system: Opportunities, barriers and policy challenges. *Clinical Pharmacology and Therapeutics* 2007; 82:427-434.
- Deverka PA, Lavalley DM, Desai PJ, et al. Stakeholder participation in comparative effectiveness research: Defining a framework for effective engagement. *Journal of Comparative Effective Research* 2012a; 1(2): 181-194.
- Deverka PA, Schully SD, Ishibe N, et al. Stakeholder assessment of the evidence for cancer genomic tests: Insights from three case studies. *Genet Med* 2012b; 14(7):656-62.
- Dressler LG. Return of research results from pharmacogenomic versus disease susceptibility studies: What's drugs got to do with it? *Pharmacogenomics* 2012; 13(8):935-49.
- Eichler HG, Oye K, Baird LG, Abadie E, Brown J, et al. Adaptive Licensing: Taking the Next Step in the Evolution of Drug Approval. *Clin Pharmacol & Ther* 2012; 91(3): 426-37.
- Elkin EB, Weinstein MC, Winer EP, et al. HER-2 testing and Trastuzumab therapy for metastatic breast cancer: A cost-effectiveness analysis. *Journal of Clinical Oncology* 2004; 22(5):854-863.
- Engstrom PF, Bloom MG, Demetri GD, et al. National Comprehensive Cancer Network. NCCN molecular testing white paper: Effectiveness, efficiency, and reimbursement. *J Natl Compr Canc Netw* 2011; 9(Suppl 6):S1-16.
- Evaluation of Genomic Applications in Practice and Prevention (EGAPP). EGAPP Working Group: Recommendation Statements. *Genet Med* 2009; 11:15-73.

Febbo PG, Ladanyi M, Aldape KD, et al. NCCN Task Force Report: Evaluating the clinical utility of tumor markers in oncology. *J Natl Compr Canc Netw* 2011; 9(Suppl 5):S1-32.

Food and Drug Administration (FDA). Guidance for Industry, Enrichment Strategies for Clinical Trials to Support Approval of Human Drugs and Biologic Products. *Draft Guidance*. December 2012a. Retrieved at: <<http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM332181.pdf>>

Food and Drug Administration (FDA). "Report to Congress: Reports on Postmarketing Studies [FDAMA 130]." Accessed 20 December 2012b. 2001. Retrieved at: <<http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Post-marketingPhaseIVCommitments/ucm071529.pdf>>.

Food and Drug Administration (FDA). Investigational Device Exemptions. 21 CFR §812.8(m). 2011a.

Food and Drug Administration (FDA). FDA-CMS Parallel Review. October 2011b. Retrieved at: <<http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/HowtoMarketYourDevice/PremarketSubmissions/ucm255678.htm>>.

Food and Drug Administration (FDA). In Vitro Companion Diagnostic Devices, Draft Guidance for Industry and Food and Drug Administration Staff. 2011c. Accessed 21 December 2012. Retrieved at: <<http://www.fda.gov/downloads/MedicalDevices/.../ucm071455.pdf>>.

Food and Drug Administration (FDA). Center for Devices and Radiological Health Public Meeting. Oversight of Laboratory Developed Tests (LDTs). In; July 19-20, 2010.

Food and Drug Administration (FDA). Draft Guidance for Industry, Clinical Laboratories, and FDA Staff - In Vitro Diagnostic Multivariate Index Assays. 2007a.

Food and Drug Administration (FDA). Guidance for Industry-Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. 2007b. Retrieved at: <<http://www.fda.gov/cber/gdlns/probl.pdf>>.

Food and Drug Administration (FDA). Guidance on pharmacogenetic tests and genetic tests for heritable markers. 2007. Accessed 31 December 2012. Retrieved at: <<http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm077862.htm>>.

Food and Drug Administration (FDA). Guidance for Industry: Pharmacogenomic Data Submissions. March 2005. Accessed 20 April 2013. Retrieved at: <http://www.fda.gov/downloads/RegulatoryInformation/Guidances/ucm126957.pdf>

Freidlin B, McShane LM, Korn EL. Randomized Clinical Trials With Biomarkers: Design Issues. *Journal of the National Cancer Institute*. 2010; 102(3):152-160.

Genoptix. Genoptix Medical Laboratory Report. Retrieved at: <www.genoptix.com>.

Gliklich RE & Dreyer NA (eds). Registries for evaluating patient outcomes: A User's Guide. (Prepared by Outcome DEClIDE Center.) AHRQ Publication No. 07-EHC001-1. Rockville, MD: Agency for Healthcare Research and Quality. April 2007. Accessed 26 December 2012. Retrieved at: <<http://www.effectivehealthcare.ahrq.gov/repFiles/PatOutcomes.pdf>>

GRACE Principles. "The GRACE Initiative: Good Research for Comparative Effectiveness." 2010. Retrieved at: <www.graceprinciples.org>.

Green ED & Guyer MS. National Human Genome Research Institute. Charting a course for genomic medicine from base pairs to bedside. *Nature* 2011; 470:204-213.

Grimwade D, Jovanoci JV, Hills RK, et al. Prospective minimal residual disease monitoring to predict relapse of acute promyelocytic leukemia and to direct pre-emptive arsenic trioxide therapy. *J Clin Oncol* 2009; 27:3650-3658.

Guilhot F, Roy L, Tomowiak C. Current treatment strategies in chronic myeloid leukemia. *Curr Opin Hematol* 2012; 19:102-109.

Haddow J & Palomaki G. ACCE: A Model Process for Evaluating Data on Emerging Genetic Tests. In: Human Genome Epidemiology: A Scientific Foundation for Using Genetic Information to Improve Health and Prevent Disease. Khoury M, Little J, Burke W (eds.) Oxford University Press. 2003; 217-233.

Helfand M, Berg A, Flum D, et al. "Draft Methodology Report: 'Our Questions, Our Decisions: Standards for Patient-centered Outcomes Research.'" Patient-Centered Outcomes Research Institute. 23 July 2012. Washington, D.C. Retrieved at: <<http://pcori.org/assets/MethodologyReport-Comment.pdf>>.

Henry NL & Hayes DF. Uses and abuses of tumor markers in the diagnosis, monitoring, and treatment of primary and metastatic breast cancer. *Oncologist* 2006;11:541-52.

Hillner BE, Smith TJ, Desch CE. Assessing the cost effectiveness of adjuvant therapies in early breast cancer using a decision analysis model. *Breast Cancer Research and Treatment* 1993; 25:97-105.

Hurley R, Jirik CL, Christensen-Johnson DL, et al. ASCO Quality Oncology Practice Initiative: Harnessing the electronic medical record to provide decision support and to improve and measure quality oncology care. *J Clin Oncol* 28, 2010 (suppl; abstr e16557). Accessed October 31, 2012. Retrieved at: <http://meetinglibrary.asco.org/content/47725-74>.

Institute of Medicine (IOM). "Demonstrating Clinical Utility in Oncology – Workshop Summary." The National Academies Press, in press. Washington, DC.

Institute of Medicine (IOM). "Evolution of Translational Omics: Lessons Learned and the Path Forward." The National Academies Press. 2012a. Washington, DC.

Institute of Medicine (IOM). "Genome-Based Diagnostics: Clarifying Pathways to Clinical Use: Workshop Report." The National Academies Press. 2012b. Washington, D.C.

Institute of Medicine (IOM). "Generating Evidence for Genomic Diagnostic Test Development:

Workshop Summary." The National Academies Press. 2011. Washington, D.C.

Institute of Medicine (IOM). "Facilitating Development and Utilization of Genome-Based Diagnostic Technologies: A Workshop." November 15, 2011. Details available at: <http://www.iom.edu/Activities/Research/GenomicBasedResearch/2011-NOV-15.aspx>.

Institute of Medicine (IOM). "Rare Diseases and Orphan Products: Accelerating Research and Development." The National Academies Press. 2010. Washington, D.C.

ISPOR Performance-based Risk-Sharing Task Force. Developing Good Research Practices for Performance-Based Risk-Sharing Arrangements. In: Forum Presentation, ed. International Society for Pharmacoeconomics and Outcomes Research (ISPOR), 14th Annual European Congress. Madrid, Spain; 2011.

Janes H, Pepe MS, Bossuyt PM, et al. Measuring the performance of markers for guiding treatment decisions. *Annals of Internal Medicine* 2011; 154:253-259.

Janssens A, Ioannidis J, van Duijn C, et al. Strengthening the reporting of genetic risk prediction studies: the GRIPS statement. *Genome Med* 2011; 3:1-5.

Javitt GH & Hudson K. Federal neglect: Regulation of genetic testing. *Issues in Science and Technology* 2006; 22:59–66.

Jennings L, Van Deerlin VM, Gully ML; the College of American Pathologists Molecular Pathology Resource Committee. Recommended principles and practices for validating clinical molecular pathology tests. *CAP Laboratory Improvement Programs. Arch Pathol Lab Med* 2009; 133:743-755.

Johnson JR, Williams G, Pazdur R. End points and United States Food and Drug Administration approval of oncology drugs. *Journal of Clinical Oncology* 2003; 21:1404-11.

Jonker DJ, O'Callaghan CJ, Karapetis CS, et al. Cetuximab for the treatment of colorectal cancer. *N Engl J Med* 2007; 357:2040-8.

Karapetis CS, Khambata-Ford S, Jonker DJ, et al. K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *N Engl J Med* 2008; 359:1757-65.

Khleif SR, Doroshow JH, Hait WN. AACR-FDA-NCI Cancer biomarkers collaborative consensus report: Advancing the use of biomarkers in cancer drug development. *Clin Can Res* 2010; 16:3299-318.

Khoury MJ, Rich EC, Randhawa G, et al. comparative effectiveness research and genomic medicine: An evolving partnership for 21st century medicine. *Genetics in medicine* 2009; 11(10):707-711.

Khoury K, Little J, Burke W (eds). "Improve Health and Prevent Disease." Oxford University Press. 2003; 217-233.

Khoury MJ, Coates RJ, Evans JP. Evidence-based classification of recommendations on use of genomic tests in clinical practice: Dealing with insufficient evidence. *Genet Med* 2010; 12:680-3.

Kühnl A & Grimwade D. Molecular Markiers in Acute Myeloid Luekaemia. *Int J Hematol* 2012; 96:153-163.

Lesko LJ, Zineh I, Huang SM. What is clinical utility and why should we care? *Clin Pharmacol Ther* 2010a; 88.

Lesko LJ, Zineh I. DNA, drugs and chariots: On a decade of pharmacogenomics at the FDA. *Pharmacogenomics* 2010b; 11:507-12.

Lewin Group. The value of laboratory screening and diagnostic tests for prevention and health care improvement. 2009. Accessed 20 December 2012. Retrieved at: <http://www.labresultsforlife.org/briefing/lewin_ac_la_value_of_lab_sx_dx_report.pdf>.

Lin JS, Thompson M, Goddard KAB. Evaluating genomic tests from bench to bedside: A practical framework. *BMC Medical Informatics and Decision Making* 2012; 12:117.

Lord SJ, Irwig L, Bossuyt PM. Using principles of randomized controlled trial design to guide test evaluation. *Med Decis Making* 2009; E1-E12.

Maitournam A & Simon R. On the efficiency of targeted clinical trials. *Stat Med* 2005; 24:329-339.

Mandrekar SJ & Sargent DJ. Predictive biomarker validation in practice: lessons from real trials. *Clin Trials* 2010; 7:567-73.

Mandrekar SJ & Sargent DJ. Clinical trial designs for predictive biomarker validation: One size does not fit all. *J Biopharm Stat* 2009a; 19:530-42.

Mandrekar SJ & Sargent DJ. Clinical trial designs for predictive biomarker validation: Theoretical considerations and practical challenges. *J Clin Oncol* 2009b; 27:4027-34.

Marchiò C, Dowsett M, Reis-Filho JS. Revisiting the technical validation of tumour biomarker assays: How to open a Pandora's box. *BMC Med* 2011; 9.

Marin D, Ibrahim AR, Lucas C, et al. Assessment of BCR_ABL1 transcript levels at 3 months is the only requirement for predicting outcome for patients with chronic myeloid leukemia treated with tyrosime kinase inhibitors. *J Clin Oncol* 2011; 30:232-238.

Mattocks CJ, Morris MA, Matthijs G, et al. The EuroGentest Validation Group. A standardized framework for the validation and verification of clinical molecular genetic tests. *Eur J Hum Genet* 2010; 18:1276-88.

McDermott U, Downing JR, Stratton MR. Genomics and the continuum of cancer care. *N Engl J Med* 2011; 364:340-50.

McShane LM. Statistical challenges in the Development and evaluation of marker-based clinical tests. *BMC Medicine* 2012; 10:52.

McShane LM, Altman DG, Sauerbrei W, et al. Statistics Subcommittee of NCI-EORTC Working Group on Cancer Diagnostics: REporting recommendations for tumor MARKer prognostic studies (REMARK). *Nat Clin Pract Urol* 2005; 2:416-22.

McShane LM & Hayes DF. Publication of tumor marker research results: the necessity for complete and transparent reporting. *J Clin Onc* 2012; 30(34):4223-4232.

McShane LM & Polley MYC. Development of omic-based clinical tests: The challenge of achieving statistical robustness and clinical utility (submitted for publication) 2013.

Meshinchi S, Hunger SP, Aplenc R et al. Lessons learned from the Investigational Device Exemption review of Children's Oncology Group Trial AML 1031. *Clin Cancer Res* 2012; 18:1547-1554.

Messner DA, Tunis SR. Current and future state of FDA-CMS parallel reviews. *Clinical Pharmacology & Therapeutics* 2012; 91:383-5.

Methods Guide for Medical Test Reviews. AHRQ Publication No. 12-EC017. Rockville, MD: Agency for Healthcare Research and Quality. June 2012. Also published as a special supplement to the *Journal of General Internal Medicine*. July 2012. Retrieved at: <<http://www.effectivehealthcare.ahrq.gov/reports/final.cfm>>.

Micheel CM, Nass SJ, Omenn GS (eds.) "Evolution of Translational Omics: Lessons Learned and the Path Forward." 2012. National Academies Press; Washington, D.C.

Moons KGM, Altman DG, Vergouwe Y, et al. Prognosis and prognostic research: Application and impact of prognostic models in clinical practice. *BMJ* 2009a; 338:b606.

Moons KGM, Royston P, Vergouwe Y, et al. Prognosis and prognostic research: What, why, and how? *BMJ* 2009b; 338:b375.

Moore HM, Kelly AB, Jewell SD, et al. Biospecimen Reporting for Improved Study Quality (BRISQ). *Cancer Cytopathol* 2011; 119:92-101.

National Cancer Institute (NCI). NCI Dictionary of Cancer Terms. Accessed 20 December 2012. 2012. Retrieved at: <<http://www.cancer.gov/dictionary?Cdrid=618612>>.

National Cancer Institute (NCI). The Nation's Investment in Cancer Research: A Plan and Budget Proposal for Fiscal Year 2008. NIH Publication No.

06-6090. 2008. Retrieved at: <http://plan.cancer.gov/plan_overview.shtml>.

National Committee on Vital and Health Statistics. "Frequently Asked Questions about Medical and Public Health Registries." Health and Human Services. Accessed 20 December 2012. Retrieved at: <<http://ncvhs.hhs.gov/9701138b.htm>>.

National Human Genome Research Institute. The National Institutes of Health Glossary of Talking Genetic Terms. Accessed 20 December 2012. 2012. Retrieved at: <<http://www.genome.gov/glossary/index.cfm?id=134>>.

National Institutes of Health (NIH). Genetics Glossary Reference. Accessed 13 January 2011. Retrieved at: <http://ghr.nlm.nih.gov/glossary_>.

National Institutes of Health (NIH). Patient-Reported Outcomes Measurement Information System: Dynamic tools to measure health outcomes from the patient perspective. 2007. Retrieved at: <<http://www.nihPROMIS.org>>.

Nguyen KH, Kobayashi S, Costa DB. Acquired resistance to epidermal growth factor receptor tyrosine kinase inhibitors in non-small-cell lung cancers dependent on the epidermal growth factor receptor pathway. *Clinical Lung Cancer* 2009; 10:281-9.

Ong FS, Das K, Wang J, et al. Personalized medicine and pharmacogenetic biomarkers: Progress in molecular oncology testing. *Expert Re Mol Diagn* 2012; 12(6):593-602.

Parkinson DR, Johnson BE, Sledge GW. Making personalized cancer medicine a reality: Challenges and opportunities in the development of biomarkers and companion diagnostics. *Clin Cancer Res* 2012;18:619-624.

Patrick D, Guyatt GH, Acquadro C. Chapter 17: Patient-reported outcomes. In: Higgins JPT, Green S (eds). *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.0.1. The Cochrane Collaboration. 2008. Retrieved at: <www.cochrane-handbook.org>.

- Patterson SD, Cohen N, Karnoub M, et al. Prospective-retrospective biomarker analysis for regulatory consideration: white paper from the industry pharmacogenomics working group. *Pharmacogenomics* 2011;12(7):939-951.
- Pepe MS, Janes H, Gu JW. Letter regarding article, "Use and misuse of the receiver operating characteristic curve in risk prediction." *Circulation* 2007; 116:e132.
- Pepe MS, Janes H, Longton G, et al. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am J Epidemiol* 2004; 159:882-90.
- Pepe MS, Janes HE. Gauging the performance of SNPs, biomarkers, and clinical factors for predicting risk of breast cancer. *J Natl Cancer Inst* 2008; 100:978-9.
- Pettiti DB, Teutsch SM, Barton MB, et al. U.S. Preventative Services Task Force. Update on the methods of the U.S. preventative services task force: Insufficient evidence. *Ann Intern Med* 2009; 150:199-205.
- Plevritis SK. Decision analysis and simulation modeling for evaluating diagnostic tests on the basis of diagnostic outcomes. *AJR* 2005; 185:581-590.
- Polley MYC, Freidlin B, Korn EL, et al. Statistical and Practical Considerations for Clinical Evaluation of Predictive Biomarkers (JNCI, in press), 2013.
- Poste G, Carbone DP, Parkinson DR, et al. Leveling the playing field: Bringing development of biomarkers and molecular diagnostics up to the standards for drug development. *Clin Cancer Res* 2012; 18:1515-1523.
- Prowell TM & Pazdur R. Pathological complete response and accelerated approval in early breast cancer. *NEJM* 2012;366(26):2438-41.
- Pusztai L, Ayers M, Stec J, Hortobagyi GN. Clinical application of cDNA microarrays in oncology. *Oncologist* 2003; 8:252-8.
- Ramsey SD, Barlow WE, Gonzalez-Angulo AM, Tunis S, Baker L, Crowley J, Deverka P, Veenstra D, Hortobagyi GN. Integrating comparative effectiveness design elements and endpoints into a Phase III, randomized clinical trial (SWOG S1007) evaluating Oncotype Dx-guided management for women with breast cancer involving lymph nodes. *Contemp Clin Trials* 2013; 34:1-9 (Published online September 18, 2012).
- Ransohoff DF. Rules of evidence for cancer molecular-marker discovery and validation. *Nat Rev Cancer* 2004; 4:309-14.
- Reitsma JB, Rutjes AWS, Whiting P, et al. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0. In: Deeks JJ BP, Gatsonis C (eds). *The Cochrane Collaboration*; 2009.
- Roberts M, Russel LB, Paltiel D. Conceptualizing a model: A report of the ISPOR-SMDM modeling good research practices task force-2. *Value in Health* 2012; 15:804-811.
- Robson ME, Storm CD, Weitzel J, et al. American Society of Clinical Oncology policy statement update: Genetic and genomic testing for cancer susceptibility. *J Clin Oncol* 2010; 28:893-901.
- Rogowski WH, Grosse SD, Khoury MJ. Challenges of translating genetic tests into clinical and public health practice. *Nat Rev Genet* 2009; 10:489-95.
- Rosenberg L, Joseph, L, Barkun A. "Surgical Arithmetic: Epidemiological, Statistical and Outcome-Based Approach to Surgical Practice." 2000. Landes Bioscience; Georgetown, TX.
- Royston P, Parmar MKB, Altman DG. Visualizing length of survival in time-to-event studies: A compliment to Kaplan-Meier plots. *JNCI* 2008; 100: 92-97.
- Royston P, Moons KGM, Altman DG et al. Prognosis and prognostic research: Developing a prognostic model. *BMJ* 2009; 338:b604.
- Royston P, Parmar MKB, Altman DG. Visualizing length of survival in time-to-event studies: A complement to Kaplan-Meier plots. *JNCI* 2008; 100(2):92-97.
- Rubin DB. The design versus the analysis of observational studies for causal effects: Parallels with the design of randomized trials. *Stat Med* 2007; 26:20-36.

Saijo N. Critical comments for roles of biomarkers in the diagnosis and treatment of cancer. *Cancer Treat Rev* 2012; 38:63-7.

Salto-Tellez M, Koay ESC. Molecular diagnostic cytopathology: Definitions, scope and clinical utility. *Cytopathology* 2004; 15(5):252-255.

Samson D & Schoelles KM. Developing the topic and structuring systematic reviews of medical tests: utility of PICOTS, analytic frameworks, decision trees, and other frameworks. AHRQ Publication No. 12-EHC073-EF. Chapter 2 of *Methods Guide for Medical Test Reviews* (AHRQ Publication No. 12-EHC017). Rockville, MD: Agency for Healthcare Research and Quality; June 2012. www.effectivehealthcare.ahrq.gov/reports/final.cfm. Also published in a special supplement to the *Journal of General Internal Medicine: J Gen Intern Med* 2012; June(27 Suppl 1):S11-9.

Scheuner MT, Hilborne L, Brown J et al. A report template for molecular genetic tests designed to improve communication between the clinician and laboratory. *Genetic Testing and Molecular Biomarkers* 2012; 16:761-769.

Schilsky RL, Doroshow JH, LeBlanc M, et al. Development and use of integral assays in clinical trials. *Clin Cancer Res* 2012; 18:1540-1546.

Schulman KA & Tunis SR. A policy approach to the development of molecular diagnostic tests. *Nat Biotechnol* 2010; 28:1157-9.

–Secretary’s Advisory Committee on Genetics, Health, and Society (SACGHS). Realizing the potential of pharmacogenomics: opportunities and challenges. Report of the SACGHS. Bethesda, MD: Department of Health and Human Services, May 2008b. Accessed 1 October 2012. Retrieved at: http://www4.od.nih.gov/oba/SACGHS/reports/SAC-GHS_PGx_Report.pdf.

Secretary’s Advisory Committee on Genetics, Health, and Society (SACGHS). U.S. System of Oversight of Genetic Testing: A Response to the Charge of the Secretary of Health and Human Services, Bethesda, MD: Department of Health and Human Services, April 2008b. Accessed 1 October 2012. Retrieved at: http://oba.od.nih.gov/oba/SACGHS/reports/SACGHS_oversight_report.pdf

–Secretary’s Advisory Committee on Genetic Testing (SACGT). Enhancing the oversight of genetic tests: Recommendations of the SACGT, Secretary’s Advisory Committee on Genetic Testing. Bethesda, MD: National Institutes of Health, June 2000. Accessed 21 December 2012. Retrieved at: http://www4.od.nih.gov/oba/sacgt/reports/oversight_report.pdf.

Sepucha KR, Belkora J. Putting shared decision-making to work in breast and prostate cancers: Tools for community oncologists. *Commun Oncol* 2007;4:685-689,691.

Simon R, Maitournam A. Evaluating the efficiency of targeted designs for randomized clinical trials. *Clin Cancer Res* 2004; 10:6759-63.

Simon R, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J Natl Cancer Inst* 2009; 101:1446-52.

Simon R. Clinical trial designs for evaluating the medical utility of prognostic and predictive biomarkers in oncology. *Personalized Medicine* 2010; 7:33-47.

Simon R. Development and validation of therapeutically relevant multi-gene biomarker classifiers. *J Natl Cancer Institute* 2005a; 97:866-867.

Simon R. Roadmap for developing and validating therapeutically relevant genomic classifiers. *J Clin Oncol* 2005b; 23:7332-41.

Simon R. The use of genomics in clinical trial design. *Clin Cancer Res* 2008; 14:5984-93.

Simon R. Transforming correlative science to predictive personalized medicine. Accessed 27 March 2012. In: ASCO. Retrieved at: <http://linus.nci.nih.gov/techreport/ASCO-06-Predictive-B.ppt>.

Sox HC, Greenfield S. Comparative effectiveness research: A report from the Institute of Medicine. *Ann Intern Med*. 2009; 151(3):203-5.

Steve G, Kay D, Renee W. Gene Expression Profile Tests for Early Stage Breast Cancer. Effectiveness Guidance Document: Methodological Guidance for the Design of Comparative Effectiveness Studies. 2009. Center for Medical Technology Policy. Baltimore, Maryland. Retrieved at: <<http://www.cmtpn.org/effectiveness-guidance-documents/gene-expression-predictors-for-breast-cancer-egd/>>.

Sun F, Bruening W, Uhl S., et al. Quality, Regulation and Clinical Utility of Laboratory-developed Molecular Tests: Technology Assessment Report (Project ID: LABC0707). Rockville, MD: Agency for Health Research and Quality, Original date: May 19, 2010; Correction date: October 6, 2010. Accessed 1 October 2012. Retrieved at: <http://www.cms.gov/Medicare/Coverage/DeterminationProcess/downloads/id72TA.pdf>

Sun F, Bruening W, Erinoff E, Schoelles KM. Addressing Challenges in Genetic Test Evaluation. Evaluation Frameworks and Assessment of Analytic Validity. Methods Research Report (Prepared by the ECRI Institute Evidence-based Practice Center under Contract No. HSA 290-2007-10063-I.) AHRQ Publication No. 11-EHC048-EF. Rockville, MD: Agency for Healthcare Research and Quality. June 2011. Accessed 4 March 2013. Retrieved at at: www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Taube S, Clarke G, Dancey J, McShane L, Sigman C, Gutman S. A perspective on challenges and issues in biomarker development and drug and biomarker codevelopment. . J Natl Cancer Inst 2009; 101:1453-63.

Teutsch SM, Bradley LA, Palomaki GE, et al. EGAPP Working Group. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative: Methods of the EGAPP Working Group. Genet Med. 2009; 11(1):3-14.

Tevak, Z, Kondratovich M, Mansfield E. US FDA and Personalized Medicine: In vitro diagnostic regulatory perspective. Personalized Medicine 2010; 7(5):517–530.

Thariani R, Wong W, Carlson JJ, et al. Prioritization in Comparative Effectiveness Research: The CANCERGEN Experience. Med Care 2012; 50(5):388-93.

Trikalinos TA, Kulasingam S, Lawrence WF. Chapter 10: Deciding whether to complement a systematic review of medical tests with decision modeling. J Gen Intern Med 2012; 27(Supple 1):S76-82.

Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: Increasing the value of clinical research for decision- making in clinical and health policy. JAMA 2003; 290:1624-32.

UnitedHealth Center for Health Reform & Modernization. Working Paper 7: Personalized Medicine: Trends and Prospects for the New Science of Genetic Testing and Molecular Diagnostics. 2012. Retrieved at: <http://www.unitedhealthgroup.com/hrm/UNH_WorkingPaper7.pdf>.

Van Cutsem E, Peeters M, Siena S, et al. Open-label phase III trial of panitumumab plus best supportive care compared with best supportive care alone in patients with chemotherapy-refractory metastatic colorectal cancer. J Clin Oncol 2007; 25:1658-64.

Varadhan R & Seeger J. Estimation and reporting of heterogeneity of treatment effects. In: Velentags P, Dreyer NA, Nourja P, et al., eds. Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide. AHRQ Publication No. 12(13)-EHC099. Rockville, MD: Agency for Healthcare Research and Quality; January 2013: Chapter 3, pp. 35-44. Retrieved at: <http://www.effectivehealthcare.ahrq.gov/ehc/products/440/1166/User-Guide-to-Observational-CER-1-10-13.pdf>.

Veenstra DL, Piper M, Haddow JE, et al. Improving the efficiency and relevance of evidence-based recommendations in the era of whole-genome sequencing: an EGAPP methods update. Genet Med. 2013;15:14-24.

Velentgas P, Dreyer NA, Nourjah P, Smith SR, Torchia MM, eds. Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide. AHRQ Publication No. 12(13)-EHC099. Rockville, MD: Agency for Healthcare Research and Quality; January 2013. Retrieved at: www.effectivehealthcare.ahrq.gov/Methods-OCER.cfm.

Whiting PF, Rutjes AW, Westwood ME, et al. The QUADAS-2 Group. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; 155:529-36.

Wittes J. Times to event: Why are they hard to visualize? *JNCI* 2008; 100(2):80-81.

Wideroff L, Phillips KA, Randhawa G, et al. A health services research agenda for cellular, molecular and genomic technologies in cancer care. *Public Health Genomics* 2009; 12:233-244.

Williams PM, Lively TG, Jessup JM et al. Bridging the gap: Moving predictive and prognostic assays from research to clinical use. *Clinical Cancer Research* 2012; 18:1531-1539.

Wilson C, Schulz S, Waldman SA. Biomarker development, commercialization, and regulation: Individualization of medicine lost in translation. *Clinical Pharmacology & Therapeutics* 2007; 81:153-5.

Woodcock J. Assessing the clinical utility of diagnostics used in drug therapy. *Clin Pharmacol Ther* 2010; 88:765-73.

Yang M, Rajan S, Issa AM. Cost effectiveness of gene expression profiling for early stage breast cancer: A decision-analytic model. *Cancer* 2012; Oct15:5163-5170.

Zujewski JA & Kamin L. Trial assessing individualized options for treatment for breast cancer: the TAILORx trial. *Future Oncol* 2008; 4:603-10.

APPENDIX A: KEY INFORMANTS

Stakeholder Name	Stakeholder Category	Organizational Affiliation
Becker, Robert	Regulators	Food and Drug Administration
Bradley, Linda	Researchers	Women's & Children's Hospital of Rhode Island
Burke, Wylie	Researcher	University of Washington
Calonge, Ned	Policymakers	Colorado Trust
Epstein, Rob	Payers & Purchasers	Independent Consultant (Formerly at Medco Research Institute)
Freedman, Andy	Researchers	National Cancer Institute
Goddard, Katrina	Researchers	Kaiser
Gorman, Mark	Patients & Consumers	National Coalition for Cancer Survivorship
Hayes, Dan	Clinician	University of Michigan
Khoury, Muin	Policymakers	Centers for Disease Control and Prevention
Lyman, Gary	Researchers	Duke University, School of Medicine
McLeod, Howard	Researchers	University of North Carolina, Chapel Hill
Miller, Amy	Industry	Personalized Medicine Coalition
Nelson, David	Industry	Epic Sciences, Inc.
O'Leary, James	Patients & Consumers	Genetic Alliance
Piper, Margaret	Payers & Purchasers	Blue Cross Blue Shield
Simon, Richard	Researchers	National Cancer Institute
Smith, Mary Lou	Patients & Consumers	Research Advocacy Network
Terry, Sharon	Patients & Consumers	Genetic Alliance
Teutsch, Steven	Policymakers	LA Dept. of Health
Whittemore, Vicky	Patients & Consumers	
Parkinson, David	Industry	New Enterprise Associates
Jacques, Louis	Payers & Purchasers	Centers for Medicare and Medicaid Services

APPENDIX B: GLOSSARY OF TERMS

Term	Definition
Absolute Risk	A measure of the risk of a certain event happening. In cancer research, it is the likelihood that a person who is free of a specific type of cancer at a given age will develop that cancer over a specified time-period (NCI Dictionary of Terms, 2012).
ACCE Framework	The Analytic Validity, Clinical Validity, Clinical Utility and Ethical, legal and social implications (ACCE) Framework provides a model process of collecting, evaluating, interpreting, and reporting data about genetic testing (Khoury et al, 2003; Haddow & Palomaki, 2003).
Actionable Tests	For the purposes of this EGD, actionable tests are those that can lead to changes in clinical management of patients (Dressler, 2012). Explicitly, the term refers to tests that predict survival or other clinical endpoints independently of any specific treatment; tests that predict response to treatment; tests that assess response to treatment; and tests that are used to identify the risk of organ-based toxicities or altered metabolism and/or response to cancer drugs (pharmacogenomic tests), as long as the test result leads to some type of clinical intervention (CMTP).
Adaptive Licensing	An innovative proposal for medical product regulation. Envisions staged market entry for medical products through stepwise learning under conditions of acknowledged uncertainty, with iterative phases of data-gathering and regulatory evaluation (Eichler et al., 2012).
Analytic Validity	Ability to accurately and reliably measure the genotype (or analyte) of interest in the clinical laboratory, and in specimens' representative of the population of interest. Analytic validity includes analytic sensitivity (detection rate), analytic specificity (1-false positive rate), reliability (e.g., repeatability of test results), and assay robustness (Teutsch et al, 2009)
Association	A relationship. In research studies, association means that two characteristics (i.e. variables or factors) are related so that if one changes, the other changes in a predictable way. An association does not necessarily mean that one variable causes the other (AHRQ Glossary of Terms, 2013).
Bias	Any factor, recognized or not, that distorts the findings of a

	study. Bias can influence the observations, results, and conclusions of the study and make them less accurate or believable (AHRQ Glossary of Terms, 2013).
Biomarker	A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention (Atkinson et al.,2001).
Chain of Evidence	Identification and use of a series of separately performed studies not originally designed to answer the clinical utility question at hand, but which individually address different components of the evidence necessary to establish that the use of an MDx test in a specific clinical situation leads to beneficial health outcomes (i.e., clinical utility). Stands in contrast to direct evidence, where a study or studies have been designed and conducted specifically to address questions of clinical utility (Teutsch et al, 2009).
Clinical Endpoint	A characteristic or variable that reflects how a patient feels, functions or survives. Clinical endpoints are distinct measurements or analyses of disease characteristics observed in a study or clinical trial that reflect the effect of a therapeutic intervention (Atkinson et al. 2001).
Clinical Laboratory Improvement Amendments (CLIA)	The Clinical Laboratory Improvement Amendments (CLIA) was passed by Congress in 1988, establishing that quality standards for all laboratory testing ensures the accuracy, reliability and timeliness of patient test results regardless of where the test was performed (FDA, 2011).
Clinical Sensitivity	The proportion of persons with a disease phenotype who test positive (SACGT, 2000).
Clinical Specificity	The proportion of persons without a disease phenotype who test negative (SACGT, 2000).
Clinical Utility	The clinical utility of a genetic test refers to the evidence of improved measurable clinical outcomes, its usefulness and added value to patient management decision-making compared with current management without genetic-testing (Teutsch et al, 2009).
Clinical Validity	The clinical validity of a genetic test refers to its ability to accurately and reliably predict the clinically defined disorder or phenotype of interest. Clinical validity encompasses clinical sensitivity and specificity (integrating analytic validity), and predictive values of positive and negative tests that take into account the disorder prevalence (the proportion of

	individuals in the selected setting who have, or will develop, the phenotype/clinical disorder of interest) (Teutsch et al, 2009).
Companion Diagnostics	A diagnostic test developed for use with a particular therapeutic product to inform treatment, including determining which patients are appropriate candidates for the therapy and tailoring decisions about medications (IOM, 2010).
Comparative Effectiveness Research (CER)	CER is the generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat and monitor a clinical condition, or to improve the delivery of care. The purpose of CER is to assist consumers, clinicians, purchasers, and policy makers to make informed decisions that will improve health care at both the individual and population levels (Sox et al. 2009).
Decision-analytic modeling	A mathematical technique used to inform clinical and policy decisions about the use of health technologies. There are two distinct components of decision-analytic modeling: conceptualization of the problem (which converts knowledge of the health care process or decision into a simplified representation of the problem), and the conceptualization of the model itself (which matches the attributes and characteristics of a particular modeling type with the needs of the problem being represented (Roberts et al, 2012).
Effectiveness	Effectiveness studies of drugs examine whether a drug or treatment works when used the way that most people take them. Effectiveness means that most people who have the disease would improve if they used the treatment (AHRQ Glossary of Terms, 2013).
Efficacy	Efficacy studies examine whether a drug or other treatment works under the best possible conditions. The study participants are carefully selected and the researchers can make sure the drug is taken properly and stored properly. The study participants may differ from other people in the general public who have the disease. A treatment that has efficacy under the best conditions may not work as well in a different group of people with the same disease (AHRQ Glossary of Terms, 2013).
Endpoint	The overall outcome that the protocol is designed to evaluate. Common endpoints are severe toxicity, disease progression, or death (NIH Glossary of Terms, 2012).

Enrichment Design Study	In an enrichment study design, the biomarker is evaluated on all patients, but random assignment is restricted to patients with specific biomarker values (i.e. biomarker positive patients) (AHRQ Glossary of Terms, 2013).
External Validity	The extent to which clinical research studies apply to broader populations. A research study has external validity if its results can be generalized to the larger population (AHRQ Glossary of Terms, 2013).
FDA-CMS Parallel Review	A pilot program for concurrent review of certain FDA premarket review submissions for medical devices with CMS national coverage determinations. (FDA, 2011).
Generalizability	See " <i>external validity</i> ."
Genotype	The genetic sequence of an individual organism, often characterized in terms of known genetic variants. This can either refer to known alleles (or types) of a single gene or to collections of genes. For example, some lung cancers have a mutant EGF receptor genotype while other lung cancers have a wild-type (or normal) EGF receptor genotype (UnitedHealth Care Center for Health Reform & Modernization, 2012).
Hazard Ratio	A measure of how often a particular event happens in one group compared to how often it happens in another group, over time. In cancer research, hazard ratios are often used in clinical trials to measure survival at any point in time in a group of patients who have been given a specific treatment compared to a control group given another treatment or a placebo. A hazard ratio of one means that there is no difference in survival between the two groups. A hazard ratio of greater than one or less than one means that survival was better in one of the groups (NCI Glossary of Terms, 2012).
Incidence	The number of new cases of a disease or condition during a defined period in a specified population, or the rate at which new events occur in a defined population. In contrast, prevalence (<i>see prevalence definition</i>) refers to all cases of a disease or condition existing in the population at a given time (NCI Glossary of Terms, 2012).
Internal Validity	The extent to which the results of a clinical research study are not biased. Several characteristics of a study affect its internal validity. Are the two groups of people being compared similar in all the important characteristics that may affect the measurements of data? Are the data collected being measured using accurate methods? (AHRQ Glossary of Terms,

	2013).
In vitro Diagnostics	In vitro diagnostic products are those reagents, instruments, and systems intended for use in diagnosis of disease or other conditions, including a determination of the state of health, in order to cure, mitigate, treat, or prevent disease or its sequelae. Such products are intended for use in the collection, preparation, and examination of specimens taken from the human body (FDA, 2011).
Laboratory-developed Molecular Tests	Laboratory-developed molecular tests, also known as homebrew or in-house molecular tests, are developed within laboratories using either FDA regulated or self-developed analyte specific reagents (ASRs) and are intended for use solely in the test developer’s laboratory (Sun, 2010).
Marker-defined Subgroups	Patient groups for a clinical study defined on the basis of biomarker expression. The expression of the biomarker (often positive or negative) can be used to determine the likely course of a disease in an untreated individual, response to therapy and level of toxicity in the different subgroups (Friedlin et al, 2010).
Molecular Diagnostics	Molecular diagnostics is the study and application of molecular biology techniques (i.e. studying molecules, such as proteins, DNA and RNA, in a tissue or fluid) and knowledge of the molecular mechanisms of disease to diagnosis, prognostication and treatment of diseases (Salto-Tellez et al, 2004).
Mutation	A change in a DNA sequence that may or may not affect a physical trait or phenotype (<i>see definition for phenotype</i>). Mutations that occur in eggs or sperm can be passed on to offspring, unlike mutations that occur in body cells (NHGRI-NIH Glossary of Terms. 2012).
Negative Predictive Value (NPV)	Indicates the likelihood that people with a negative test result would not have a condition. The higher the value of the negative predictive value (e.g. 99% would usually be considered a high value), the more useful the test is for predicting that people do not have the condition (AHRQ Glossary of Terms, 2013).
Observational Study	A type of study in which individuals are observed or certain outcomes are measured. No attempt is made to affect the outcome (NCI Glossary of Terms, 2012).
Outcome	The end result of health care practices. Examples of outcomes include: Length of life following a health care treatment (i.e.

	survival), the effects a treatment has on patients' lives (e.g. changes in his or her ability to function or changes in quality of life), undesirable events (e.g. side effects of drugs), or whether people need to change to another kind of treatment (AHRQ Glossary of Terms, 2013).
Patient-Reported Outcome (PRO)	Any reports coming directly from patients about how they function or feel in relation to a health condition and its therapy without interpretation of the patient's responses by a clinician or other person (Patrick et al., 2008).
Patient Registry	An organized system for the collection, storage, retrieval, analysis, and dissemination of information on individual persons who have: a) a particular disease, b) a condition (e.g. a risk factor) that predisposes to the occurrence of a health-related event, c) or prior exposure to substances or circumstances known or suspected to cause adverse health effects (National Committee on Vital and Health Statistics, 2012).
Pharmacogenomic test	An assay intended to study inter-individual variations in whole-genome or candidate gene, single-nucleotide polymorphism (SNP) maps, haplotype markers, or alterations in gene expression or inactivation that may be correlated with pharmacological function and therapeutic response. In some cases, the pattern or profile of change is the relevant biomarkers, rather than changes in individual markers (FDA 2005).
Phenotype	A phenotype is an individual's observable traits, such as height, eye color, and blood type (NHGRI-NIH Glossary of Terms, 2012).
Positive Predictive Value (PPV)	Indicates the likelihood that a person with a positive test result would actually have the condition for which the test is used. The higher the value of the positive predictive value (e.g. 90%), the more useful the test is for predicting that the person has the condition (AHRQ Glossary of Terms, 2013).
Pragmatic Clinical Trial (PCTs)	Studies intended to maintain the internal validity of RCTs while being designed and implemented in ways that would better address the demand for evidence about real-world risks and benefits for informing clinical and health policy decisions (Chalkidou et al, 2012).
Prevalence	The number of diagnosed cases of a particular condition or disease existing in a specified population at a given time. It is distinct from incidence, which is the number of <i>new</i> cases of

	the disease arising in the population over a given time period (NIH Genetics Glossary, 2012).
Prognostic marker	A prognostic marker can be defined as either a single trait or signature of traits that separates different populations with respect to the risk of an outcome of interest in absence of treatment, or despite non-targeted “standard” treatment (Tevak et al. 2010).
Predictive marker	A predictive marker can be defined as either a single trait or signature of traits that separate different populations with respect to the outcome of interest in response to a particular targeted treatment (Tevak et al. 2010).
Progression-free survival (PFS)	The length of time during and after treatment in which a patient is living with a disease that does not get worse. Progression-free survival may be used in a clinical study or trial to help find out how well a new treatment works (NCI Glossary of Terms, 2012).
Prospective Observational Study	An observational study for which the consequential outcomes of interest occur after study commencement, including creation of a study protocol and analysis plan, and study initiation. Patient exposure to any of the interventions being studied may have been recorded before the study initiation such as when the study uses an existing registry cohort. Exposure may include pharmaceutical intervention, surgery, medical device, prescription, and decision made to treat. This contrasts with a retrospective observational study that employs existing secondary data sources in which both exposure and outcomes have already occurred. (Berger M et al, 2012).
Prospective-Retrospective Study	Uses archived specimens from a previously conducted clinical trial with treatment(s) that are relevant to the intended clinical use of the test. This design requires a prospective written protocol describing study objectives, methods and analytical plan and the test must be locked down and analytically validated (IOM 2012).
Randomized clinical trial (RCT)	A study in which the participants are assigned by chance to separate groups that compare different treatments; neither the researchers nor the participants can choose which group. Using chance to assign people to groups’ means that the groups will be similar and that the treatments they receive can be compared objectively. At the time of the trial, it is not known which treatment is best. It is the patient’s choice to be in a randomized trial (NCI Glossary of Terms, 2012).

<p>Receiver operating curve (ROC)</p>	<p>A ROC is a graph that plots true positive rates against false positive rates for a series of cutoff values. In other words, sensitivity is plotted on the Y-axis and specificity is plotted on the X-axis for each cutoff value. The area under the ROC is the probability that a test correctly classifies patients as true positives (those who test positive for a disease actually have it) or true negatives (those who test negative for a disease do not have it). Larger areas under the ROC indicate higher accuracy (Rosenberg et al. 2000).</p>
<p>Relative Risk</p>	<p>A measure of the risk of a certain event happening in one group compared to the risk of the same event happening in another group. In cancer research, relative risk is used in prospective studies, such as cohort studies and clinical trials. A relative risk of one means there is no difference between two groups in terms of their risk of cancer, based on whether or not they were exposed to a certain substance or factor, or how they responded to two treatments being compared. A relative risk of greater than one indicates that being exposed to a certain substance or factor increases risk of cancer or treatment being compared. A relative risk less than one indicates that being exposed to a certain substance or factor decreases the risk of cancer or treatment being compared. Also often referred as risk ratio (NCI Glossary of Terms, 2012).</p>
<p>Resource Utilization</p>	<p>Resource utilization measures reflect the amount or cost of resources used to create a specific product of the health care system. These measures can be classified into three main categories including:</p> <ol style="list-style-type: none"> 1. Relatively simple measures of the resources used to produce health care, such as mean length of stay, mean charges or estimated costs, and readmission rates for hospitals; and consultation or test ordering rates for outpatients with common complaints such as low back pain. 2. More complex measures of health care resource use, including both inpatient and outpatient services, using econometric or mathematical programming techniques to account for multiple outputs. 3. Measures of the resources used in an episode of care for a patient, or to treat a patient with a specified burden of comorbidity for a specified period of time. (AHRQ Glossary of Terms, 2013).
<p>Retrospective study</p>	<p>A study that compares two groups of people: those with the disease or condition under study (cases) and a very similar</p>

	group of people who do not have the disease or condition (controls). Researchers study the medical and lifestyle histories of the people in each group to learn what factors may be associated with the disease or condition. For example, one group may have been exposed to a particular substance that the other was not. Also called case-control study (NCI Glossary of Terms, 2012).
Retrospective observational study	Employs existing secondary data sources in which both exposure and outcomes have already occurred (Berger M. et al, 2012).
Stakeholders	Individuals, organizations or communities that have a direct interest in the process and outcomes of a project, research or policy endeavor (Deverka et al. 2012a, 2012b).
Stakeholder Engagement	An iterative process of actively soliciting the knowledge, experience, judgment and values of individuals selected to represent a broad range of direct interests in a particular issue, for the dual purposes of: <ul style="list-style-type: none"> • Creating a shared understanding; and • Making relevant, transparent, and effective decisions. (Deverka et al. 2012).
Surrogate Endpoint	A biomarker that is intended to substitute for a clinical endpoint. A surrogate endpoint is expected to predict clinical benefit (or harm or lack of benefit or harm) based on epidemiological, pathophysiologic, or other scientific evidence (Atkinson et al. 2001).
Time-varying confounding	Refers to the setting in which the outcome and treatment/exposure are influenced by new values of a third variable. Disease severity changes over time, influences the decision to initiate therapy, and relates to outcome (Berger M et al., 2012).
Time-invariant confounding	Refers to a confounder that does not change values over time. For example, a patient’s socioeconomic status may be related to treatment selection and functional status, and, assuming a short observational period, does not change over time (Berger et al., 2012).