

# Generating and evaluating evidence of the clinical utility of molecular diagnostic tests in oncology

Patricia Deverka, MD, MBE<sup>1</sup>, Donna A. Messner, PhD<sup>1</sup>, Robert McCormack, PhD<sup>2</sup>, Gary H. Lyman, MD, MPH<sup>3</sup>, Margaret Piper, MD, MPH<sup>4</sup>, Linda Bradley, PhD<sup>5</sup>, David Parkinson, MD<sup>6</sup>, David Nelson, PhD<sup>7</sup>, Mary Lou Smith, JD, MBA<sup>8</sup>, Louis Jacques, MD<sup>9</sup>, Tania Dutta, MS, MPP<sup>1</sup> and Sean R. Tunis, MD, MSc<sup>1</sup>

**Purpose:** Enthusiasm for molecular diagnostic (MDx) testing in oncology is constrained by the gaps in required evidence regarding its impact on patient outcomes (clinical utility (CU)). This effectiveness guidance document proposes recommendations for the design and evaluation of studies intended to reflect the evidence expectations of payers, while also reflecting information needs of patients and clinicians.

**Methods:** Our process included literature reviews and key informant interviews followed by iterative virtual and in-person consultation with an expert technical working group and an advisory group comprising life-sciences industry experts, public and private payers, patients, clinicians, regulators, researchers, and other stakeholders.

**Results:** Treatment decisions in oncology represent high-risk clinical decision making, and therefore the recommendations give preference to randomized controlled trials (RCTs) for demonstrating CU.

The guidance also describes circumstances under which alternatives to RCTs could be considered, specifying conditions under which test developers could use prospective-retrospective studies with banked biospecimens, single-arm studies, prospective observational studies, or decision-analytic modeling techniques that make a reasonable case for CU.

**Conclusion:** Using a process driven by multiple stakeholders, we developed a common framework for designing and evaluating studies of the clinical validity and CU of MDx tests, achieving a balance between internal validity of the studies and the relevance, feasibility, and timeliness of generating the desired evidence.

*Genet Med* advance online publication 3 December 2015

**Key Words:** clinical utility; comparative effectiveness research; molecular diagnostic test; patient-reported outcome; stakeholder engagement

## INTRODUCTION

The Center for Medical Technology Policy convenes stakeholders to develop effectiveness guidance documents (EGDs), which provide disease- or technology-specific methodological recommendations for studies targeting the information needs of payers, with input from clinicians and patients. EGDs are analogous and complementary to US Food and Drug Administration (FDA) regulatory guidance documents, focusing on study designs that address payers' expectations for evidence.

Groups conducting technology assessments or systematic evidence reviews, or translating evidence into clinical practice guidelines, have frequently concluded that the evidence supporting the clinical use of a recently introduced molecular diagnostic (MDx) test is insufficient.<sup>1</sup> While these assessments typically include identification of the critical gaps in knowledge that limit the translation of specific tests into practice, they often stop short of providing specific guidance for study design to overcome these deficiencies, nor do they provide test

developers with a clear sense of the evidence that public and private health plans require for coverage.<sup>2</sup> Yet relatively few commercially available MDx tests are reviewed for coverage because of a lack of clinical utility (CU) studies.<sup>3</sup> This EGD<sup>4</sup> provides to test developers specific recommendations to evaluate the clinical validity (CV) and CU of "actionable" MDx tests in a manner that is acceptable to payers, and it serves as a resource for payers to communicate standards of evidence to test developers.

We used "molecular diagnostic test for oncology" as an umbrella term for any test that, at the molecular level, helps to identify patients with an inherited risk for cancer or to diagnose, classify, or guide management of a patient's cancer. This definition included tests for individual biomarkers, "omics"-based tests, and tests for circulating tumor cells; was independent of the assay method; and applied to tests that were *not* codeveloped as companion diagnostics. Codeveloped companion diagnostics were excluded from the scope because these test-drug combinations undergo FDA review, and this process typically results in adequate information regarding the utility

<sup>1</sup>Center for Medical Technology Policy, Baltimore, Maryland, USA; <sup>2</sup>Janssen Oncology, Johnson & Johnson, Raritan, New Jersey, USA; <sup>3</sup>Division of Medical Oncology, Fred Hutchinson Cancer Research Center, University of Washington, Seattle, Washington, USA; <sup>4</sup>Center for Health Research, Kaiser Permanente Northwest, Portland, Oregon, USA; <sup>5</sup>Department of Pathology and Laboratory Medicine, Alpert Medical School, Brown University, Providence, Rhode Island, USA; <sup>6</sup>New Enterprise Associates, Inc., Menlo Park, California, USA; <sup>7</sup>Epic Sciences, San Diego, California, USA; <sup>8</sup>Research Advocacy Network, Chicago, Illinois, USA; <sup>9</sup>ADVI, Washington, DC, USA. Correspondence: Donna A. Messner ([donna.messner@cmtpn.org](mailto:donna.messner@cmtpn.org))

Submitted 24 February 2015; accepted 30 September 2015; advance online publication 3 December 2015. doi:10.1038/gim.2015.162

of the test for the approved indication. Most other tests do not, and many are marketed as laboratory-developed tests that are regulated under the CLIA of 1988. We used the ACCE framework (analytic validity (AV), CV, CU, and ethical, legal, and social implications) to categorize the types of evidence needed to recommend the use of MDx tests.<sup>5</sup>

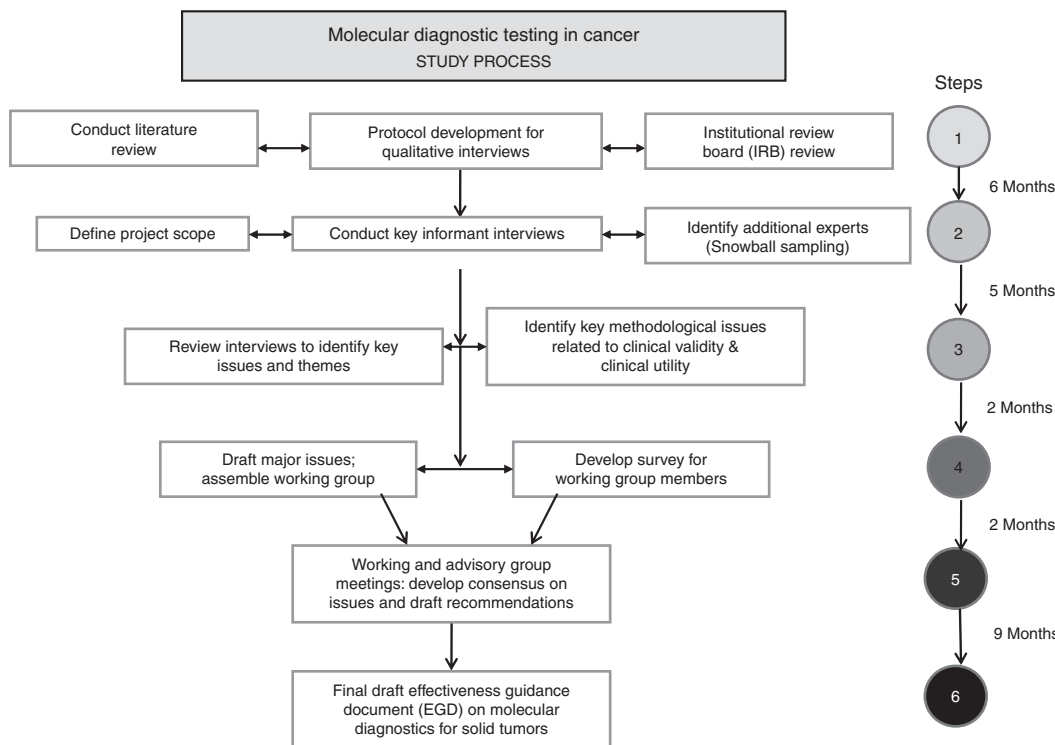
The recommendations apply to *actionable* tests, meaning tests that can lead to changes in the clinical management of patients, predict survival or other clinical end points independent of any specific treatment (“prognostic test”), predict response to treatment (“therapy-guiding test” or “predictive test”), and assess response to treatment (“monitoring test”), and that are used to identify the risk of organ-based toxicities or altered metabolism and/or response to cancer drugs (“pharmacogenomic test”). The target condition can involve either solid or hematologic malignancies in adult patients. Since these tests guide patient care decisions for a potentially life-threatening clinical condition, all are classified as “high risk” in terms of the potential benefits and harms to patients.

**MATERIALS AND METHODS**

**Figure 1** outlines the process for the development of EGDs from gap identification to final EGD recommendations.<sup>4</sup>

We convened a 10-person technical working group (TWG), comprising clinical and methodological experts representing the Centers for Medicare and Medicaid Services, Blue Cross Blue Shield Technology Evaluation Center, the National Cancer Institute, Brigham and Women’s Hospital, the Effectiveness in

Genomic Application in Practice and Prevention initiative, Duke University School of Medicine, the American Society of Clinical Oncology, Veridex, Epic Sciences, New Enterprise Associates, and the Research Advocacy Network (for details, see the EGD<sup>3</sup>). The group held an initial all-day, in-person meeting followed by a series of five teleconferences over 8 months to develop draft methodological recommendations. Following those steps, a 20-person advisory group comprised of life-sciences industry experts was convened to review and comment on the draft recommendations. Two joint advisory group/TWG in-person workshops were held, with the additional participation of patients, payers, clinicians, regulators, professional societies, and researchers. Major health plans (WellPoint (now Anthem–Kaiser Permanente), UnitedHealth, Centers for Medicare and Medicaid Services, Palmetto GBA, and Blue Cross Blue Shield Technology Evaluation Center) supported the project through funding or direct participation. Between these two workshops, a series of six joint advisory group/TWG subgroups refined and finalized specific recommendations over a 5-month period. The resulting recommendations incorporate collective stakeholder input while representing standards that are acceptable to many payers for decision making regarding coverage. Effort was made to mediate conflicting opinions within the TWG, but full consensus was not achieved. The Center for Medical Technology Policy takes responsibility for the final content. Only TWG members listed as coauthors can be considered to endorse the recommendations.



**Figure 1 Overall method for developing recommendations.**

## RESULTS

Ten specific EGD recommendations are discussed here. The recommendations are divided into three categories: reporting AV, CV, and CU. Several position statements are included to emphasize the broader need to promote evidence generation.

### Reporting AV

**Recommendation 1.** Follow standard reporting guidelines to document that analytic validity has been established. Greater transparency will enable others to more easily assess these claims.<sup>6-9</sup> Although specific methodological recommendations related to AV were excluded from the scope of this guidance document, ensuring AV before the final assessment of CV is critical to improving the evidence base for MDx tests in oncology.

### Clinical validity

The strength of the association between the test result and the clinical condition of interest must be established to assess the CV of an MDx test. The most common flaws in MDx clinical validation studies include relying on intermediate outcomes that are not predictive of the definitive clinical end point of interest (e.g., progression-free survival is often not predictive of overall survival) and use populations that are not representative of the population in which the test is intended to be used (e.g., test validation with a largely Caucasian population when the underlying disease also affects large numbers of African Americans).<sup>10-12</sup> Best practices can be achieved through attention to study design and quality (i.e., bias), sample size, patient population, choice of outcome measures, and appropriate statistical analysis and result interpretation.<sup>13-15</sup>

**Recommendation 2.** Specify the clinical context and patient population intended to benefit from the action or decision guided by the test result. One or more specific intended uses for the MDx test and outcomes of interest should also be determined as early as possible in the development process.<sup>16</sup> While preliminary or exploratory studies early in test development (including the development of classifier models) might use convenience samples obtained from less representative patient subgroups, 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 should ensure that the test sets used for validation are drawn from the intended use population and are independent of any training data sets used to develop the test.

**Recommendation 3.** Report the strength of an association between the MDx test and a specific disease state using metrics that are most useful to clinicians. When the clinical disease state is binary (e.g., a continuous variable with an actionable threshold), preferred metrics are clinical sensitivity, clinical specificity, positive predictive value, and negative predictive value, provided with measures of uncertainty such as 95%

confidence intervals. Disease prevalence among the tested population is required to compute the positive predictive value and negative predictive value. The acceptable balance of false-positive versus false-negative results depends on the clinical context. Although the area under the receiver-operator characteristic curve should not be the only metric used to evaluate CV, the optimal cut point for clinical decision making can be selected using a receiver-operator characteristic curve to plot sensitivity and (1 – specificity) pairs versus the associated levels of the MDx biomarker.<sup>17,18</sup>

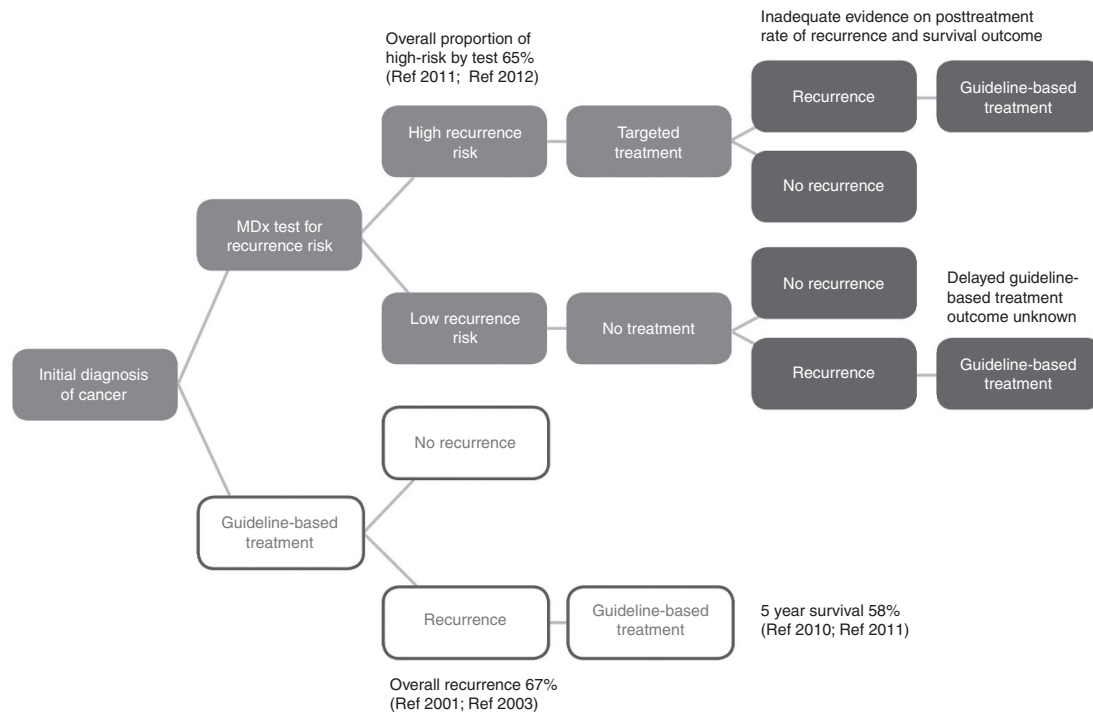
Prognostic biomarkers are typically evaluated as part of a multivariate analysis for a model predicting a particular outcome<sup>13,19</sup> and are best examined in a prospective cohort study<sup>20</sup> or possibly in the control arm of a randomized controlled trial (RCT). The preferred study design for validating a predictive biomarker is an RCT comparing two treatments, where biomarker status is available for all patients at baseline (not an enrichment design, which in this case refers to the prospective use of a patient's biomarker status for determining enrollment in a trial to increase the likelihood of observing a drug effect). When the predictive biomarker is a continuous measure, a useful approach for choosing a cutoff value is to use treatment predictiveness curves,<sup>15</sup> plotting clinical outcome (e.g., 5-year disease-free survival rate; y axis) as a function of biomarker value (x axis) separately for each treatment arm. 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 to promote reproducibility, reporting of test validation studies should utilize appropriate standards, such as the QUADAS checklist (designed to assess the quality of primary diagnostic accuracy studies)<sup>21</sup> and the REMARK (Reporting Recommendations for Tumor Marker Prognostic Studies) checklist.<sup>1,2</sup>

### Clinical utility

Evidence of the CU of an MDx test establishes the net clinical benefit to the patient of adding the MDx test to the current/standard clinical decision-making matrix. The AV and CV of the test should be “fully specified and locked down” before initiating prospective evaluations of CU.<sup>22</sup> Because these tests are used to inform oncology care decisions, they are considered high-risk medical decision tools; correspondingly high evidence standards apply. RCTs are therefore the preferred method to assess CU in this context (recommendations 4 and 5). Under specific circumstances, however, alternative study designs may be permissible (recommendations 7, 8, and 9), and in some situations, a chain of evidence might be constructed using existing evidence on therapeutics to correlate testing with patient outcomes (recommendation 10).

The earliest stages of MDx assay development should include a systematic plan for evidence-based translation into clinical practice. To determine the type(s) of studies that will be required, describe the proposed CU of the test in a flow



**Figure 2** Example flow diagram that outlines at a conceptual level the intended clinical use in practice and the associated primary patient outcomes for a clinical trial. For illustration purposes, the diagram includes some hypothetical data and reference sources for each pathway. Dark grey boxes indicate decision steps for which information does not exist or is inadequate. 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 that is still needed to fully develop the test. MDx, molecular diagnostic.

diagram (Figure 2) that outlines at a conceptual level the intended clinical use and key elements, such as the intended use population, existing test strategies, treatment alternatives, and the associated primary patient outcomes; this is analogous to defining the primary study objectives for a clinical trial.<sup>23</sup> The flow diagram serves two critical purposes: (i) helping the researcher to decide whether a prospective study is necessary by identifying existing data sources that estimate the strength of association between a test result and patient outcome(s) and (ii) helping to identify critical missing data elements, thereby supporting the design of efficient studies.

**Recommendation 4.** Specify in advance the potential therapeutic actions or decisions (i.e., clinical pathways) that should be followed based on test results, and include all relevant (for the given clinical context) treatment alternatives under consideration at the time of testing. Standardizing the potential clinical pathways associated with various test results reduces variation and enhances the ability of the study to assess the impact of test results on patient outcomes. The explicit description of how the test results will be used compared with non-biomarker-guided treatment strategies is also informative for patients who are considering enrollment in the study.

**Recommendation 5.** Include outcome measures that assess both the 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.

The primary clinical application for actionable MDx tests in oncology is to enhance the stratification of patients to more precisely classify risk and target interventions. Examples of typical outcome measures include clinical assessments of disease remission and progression, response to therapy, functional status, as well as disease- and treatment-related 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.<sup>3,24</sup> CU studies may reasonably include end points such as survival and downstream health-care resource utilization. The decision to include these end points 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 persuasive study end points unless there exists a separate, robust body of credible evidence (as determined by widely accepted evidence review standards) linking specific clinical management decisions with relevant health outcomes. Studies designed to report intended care plans following an MDx test are insufficient for demonstrating CU.

**Recommendation 6.** The preferred method for assessing the CU of MDx tests is RCTs that adequately evaluate the impact of the clinical decision (treatment or other clinical pathway)

relative to an appropriate control for both marker-positive and marker-negative patients.<sup>11</sup>

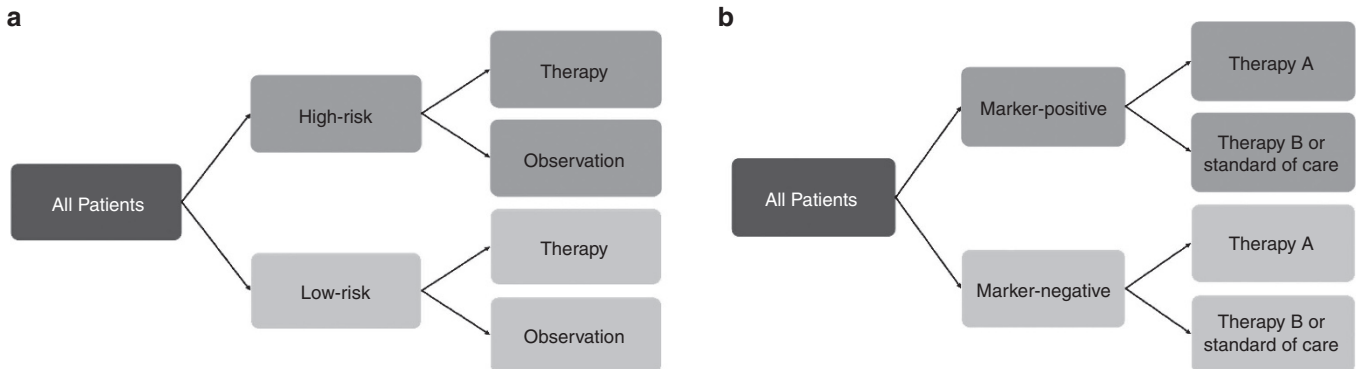
In general, designs that use a biomarker to guide the analysis are preferred over designs that use a biomarker to guide the treatment assignment.<sup>10</sup> Accordingly, a preferred RCT design is the “all comers” marker-stratified design for evaluating the CU of MDx tests<sup>4</sup> (Figure 3a,b).<sup>10-12</sup> When there exists compelling evidence that a subgroup of patients with a particular marker 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<sup>25</sup> (Figure 4a). In general, however, the approach is justified only in cases where the biologic rationale and preliminary evidence that only one group benefits is sufficiently compelling that equipoise does not truly exist between the current alternatives for all patients, making it unethical to randomize treatment options to all marker-based groups.

The biomarker strategy design, in which the patients who are randomized to usual care are not tested, is often used to study genomics-guided treatment versus usual care<sup>11</sup> (Figure 4b). With this strategy, however, some patients receiving MDx-guided therapy receive the same treatment (standard of care) as patients in the standard therapy arm, which dilutes the ability

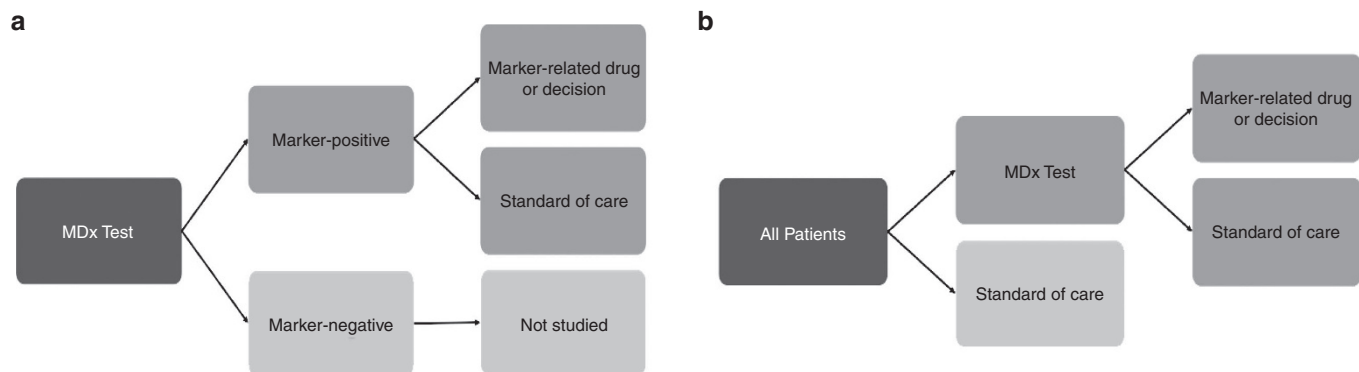
to observe a treatment effect<sup>11</sup> (Figure 4b). The same objectives can typically be achieved with fewer patients using the marker-stratified design described above. Given the larger sample size required to demonstrate a difference between study arms, the biomarker strategy design is not preferred.

**Recommendation 7.** Conduct a well-designed, prospective-retrospective study when there exists an appropriately-designed, powered, and conducted clinical trial with banked biospecimens (Figure 5). 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. To ensure the appropriate use of a “prospective-retrospective” study design to evaluate the CU of a new biomarker, several conditions must be present to ensure that this approach is of sufficient scientific rigor to convincingly demonstrate CU.<sup>26</sup> For example, the analysis plan for the biomarker study must be completely prespecified, and the analytic validity of the test must be well established to ensure that results from archived tissues resemble the results from tissue collected in real time.

Replicating validation study results is excellent verification of evidence. We believe, however, that if a single properly designed

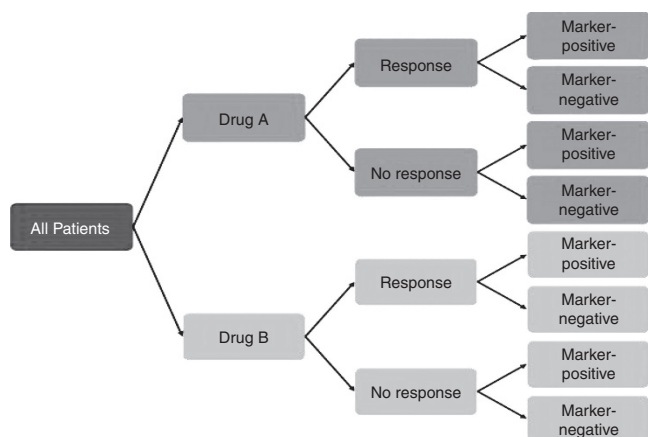


**Figure 3 Generally preferred designs.** (a) “All comers,” prospective marker-stratified design: (a) a prognostic test and (b) a predictive test. Adapted from ref. 40.



**Figure 4 Designs having disadvantages.** (a) Marker enrichment design. This is not recommended except where there exists compelling evidence 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; no information on the excluded group is obtained. (b) Biomarker strategy design. This is not preferred because the approach reduces statistical power, given that patients in both study arms receive the “standard of care” as their intervention. MDx, molecular diagnostic. Adapted from ref. 40.





**Figure 5 Prospective-retrospective randomized controlled trial (RCT) design.** A drug is tested first in an RCT, and marker-status is determined retrospectively from tissue samples. This is recommended for situations in which the marker was not known when drug was first developed. It can also be used for independent validation. Adapted from ref. 40.

and adequately powered prospective-retrospective study has positive results, this is considered adequate evidence of CU.

**Recommendation 8.** Single-arm studies can be used to establish the CU of an MDx test provided the following conditions are met: (i) the MDx test is being developed with an oncology drug that has already been approved by the FDA on the basis of pivotal trials of a study population that was not previously stratified on the basis of molecular marker status; (ii) adequate archived tissue samples are not available to conduct a prospective-retrospective trial to assess CU; (iii) it is feasible to use response, variably defined as complete or overall response, as an end point in the single-arm study; and (iv) there exists comparable response data from a noncontemporaneous comparative cohort.

This approach is applicable when an MDx test potentially identifies a subset of patients who benefit differentially from a drug treatment that has already received FDA approval on the basis of randomized trials in a broad patient population defined by disease characteristics but not biomarker status. In this setting, it would not be ethical or practicable to conduct subsequent RCTs in which a control group is denied the approved therapy. An alternative is to conduct a single-arm study. The study can be interpreted in the context of the response of a noncontemporaneous cohort or end points such as tumor shrinkage. Single-arm studies of this type are not as robust as RCTs because they provide only information on the test-positive patients, not the test-negative patients (who cannot be assumed not to benefit from the treatment). Nevertheless, marker-based differential tumor response can provide useful data to clinicians that can be used in the context of other relevant information to create an individual treatment plan.

**Recommendation 9.** Longitudinal observational study designs such as prospective cohort studies, patient registries that

explicitly include comparators, and multiple group, pretest/posttest designs (quasi-experimental) may be used as evidence of CU provided that a compelling rationale for not doing an RCT is addressed, efforts to minimize confounding factors are documented, and good research practices for prospective observational studies are followed, including public registration of studies. Since the necessary parameters for evaluating the CU 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.

The decision to pursue an observational study rather than an RCT should be considered only when other approaches are not possible; this may be particularly problematic when evaluating predictive biomarkers that compare outcomes between treatments. Factors influencing the decision include the state of clinical equipoise for the MDx test of interest and whether the proposed study design and analysis plan will sufficiently address potential problems with time-varying and time-invariant confounding and bias.<sup>27</sup> A prospective observational study should adopt best practices to minimize threats to validity. 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 to handle potential confounding, missing data, loss to follow-up, and heterogeneity of treatment effects is essential.<sup>28</sup> Various user guides on best practices for designing observational studies have been prepared by the Agency for Healthcare Research and Quality and other expert task forces, and researchers are encouraged to consult these guides before planning an observational study.<sup>27,29–31</sup>

**Recommendation 10.** Use formal decision-analytic modeling techniques to elucidate the relationship between test results, corresponding clinical pathways, and downstream patient outcomes in cases where an MDx test has established evidence of CV and plausible evidence of CU based on modeling of the initial scenario (a simplified approach for outcomes: base case, best case, worst case).

In this context, decision-analytic modeling denotes a 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.<sup>32</sup> These models are useful in the common situation where there is no direct evidence of CU. Developing a simple decision model, called a “scenario model,” that consists of a simplified decision tree and a series of “what if” scenarios can provide a quantitative assessment of the general likelihood that an MDx test will demonstrate CU. The key parameters and assumptions under three scenarios (base case, best case, and worst case) should be revisited with key stakeholders (e.g., patients, clinicians, and payers) and the outcomes estimated for each case.

For MDx tests that cross the plausibility threshold, modeling techniques are used to project the overall downstream health

outcomes (all patient-relevant benefits and harms related to the duration and quality of remaining life, such as modeled estimates of clinical events, life expectancy, and quality-adjusted life-years)<sup>33</sup> that in most instances may not be available, even within the context of RCTs, because of limited follow-up, highly selected patient populations, and/or small sample sizes. Alternatively, data from separate studies demonstrating the relationship between biomarker statuses, various steps in the care pathway, and patient outcomes may be quantitatively linked through modeling to provide estimates of the net benefit to patients.

## DISCUSSION

These recommendations aim to clarify what is adequate evidence for coverage of MDx tests. Greater clarity, consistency, and predictability of evidence requirements are essential for investors and diagnostics companies to make informed decisions regarding test development. The TWG specifically confined these recommendations to “actionable” MDx tests; they exclude tests that do not provide information leading to an alteration in clinical management. While there has been debate on the definition of “clinical utility,” our TWG rapidly came to consensus with the prevailing concept of the Effectiveness in Genomic Application in Practice and Prevention Working Group, Medicare, many evidence review groups,<sup>29,34</sup> and others<sup>35</sup> that CU refers to evidence that use of MDx test information leads to a change in patient management that can result in improved health outcomes.

This definition of *actionable* is consistent with many payers’ concept of a “medically necessary” test, which can entail consideration not only of the impact of the test on patient management but also of the current standard of care, including the adequacy of other tools available for the same purpose as the test (i.e., comparative effectiveness). The evaluation of the CU of an MDx test is, likewise, inherently a comparative effectiveness research question, requiring a comparison of the effects of the new test result versus a standard (or no) test result on patient outcomes. For this purpose, the focus is primarily on health outcomes. Health-resource utilization would also be a meaningful outcome to examine in comparative studies but was not the focus of this work, since the significance of any economic analyses is dependent on sound evidence of CV and CU.

Given the uneven quality of published studies to date, numerous groups, including the Institute of Medicine,<sup>22</sup> the National Cancer Institute,<sup>36</sup> and the National Comprehensive Cancer Network,<sup>37</sup> among others,<sup>38</sup> have published checklists, study design recommendations, and criteria for evaluating the CV and CU of MDx tests, although not always strictly limited to tests used in oncology. Our process is distinct from these in that it involved a sustained dialogue across the full range of experts and stakeholders, and emphasized the information needs and participation of major health plans. A limitation of the EGD is that it is not a consensus statement of all participants or payers generally. Nevertheless, thoughtful input of key health-plan decision makers lends confidence that tests evaluated

successfully under these guidelines can achieve affirmative coverage decisions.

Notably, the recommendations expand consideration of evidence to include not only RCTs and prospective-retrospective analyses of samples from previously conducted clinical trials but also prospective observational studies and modeling when the circumstances justify using these options. The recommendations thus reflect a growing recognition of the limitations of RCTs to address all relevant comparative questions in oncology and the usefulness of appropriately designed nonrandomized comparative effectiveness research studies.<sup>38</sup>

These recommendations create an important foundation for clarifying the evidence of CV and CU needed for coverage of MDx tests. However, as new high-throughput genomic sequencing techniques increasingly gain prominence in clinical laboratories, gradually supplanting traditional single-gene (or few gene) analyses, novel challenges arise for evaluating and covering testing. Many biomarkers originally developed as drug targets in a particular cancer can be targets in other types of cancer as well, but the effectiveness of the targeting in the new context is often unknown. How should the CU of large gene panels, or whole-exome or whole-genome sequencing, be evaluated? When is coverage appropriate? The barriers to using RCTs for assessment are all the more acute as the number of new variants to be evaluated increases. Answering these questions through a multistakeholder dialogue that includes payers—work that is underway<sup>39</sup>—is a critical next step to building constructively on the principles established in this EGD and ensuring patient access to high-quality, efficacious genomic testing for oncology decision making.

## ACKNOWLEDGMENTS

Support for this effectiveness guidance document was provided by the WellPoint Foundation, Kaiser Permanente National Community Benefit Fund at the East Bay Community Foundation, United Health Foundation, AstraZeneca, Sanofi-Aventis, Millennium: A Takeda Oncology Company, GlaxoSmithKline, Becton Dickinson, Daiichi Sankyo, Cepheid, Genentech, Amgen, Genomic Health, Myriad Genetics, and PricewaterhouseCoopers.

The authors acknowledge the time and technical expertise contributed to the technical working group by Richard Simon and Howard McLeod.

## DISCLOSURE

After this paper was written, L.J. (ADVI, Washington, DC) consulted for Myriad, GenomeDx, and Exact Sciences. It should be understood that this consultancy did not overlap with the writing of the manuscript. L.J. received an honorarium for his consultancy with Exact Sciences. After the writing of the manuscript was complete, L.J. gave expert testimony to the House Energy & Commerce Committee. D.N. discloses that he holds the position of director at the Clarity Foundation. D.N. has been compensated by Life Science Group for consultancy/advisory work and owns stock in Epic Sciences. R.T.M. discloses his position as head of technology strategy and innovation in research and development at Janssen

Oncology. R.T.M. also owns stock in Johnson and Johnson. The other authors declare no conflict of interest.

## REFERENCES

1. Khoury MJ, Berg A, Coates R, Evans J, Teutsch SM, Bradley LA. The evidence dilemma in genomic medicine. *Health Aff (Millwood)* 2008;27:1600–1611.
2. 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;11:3–14.
3. Hresko A, Haga SB. Insurance coverage policies for personalized medicine. *J Pers Med* 2012;2:201–216.
4. Deverka PA, Messner DA, Dutta T. Evaluation of clinical validity and clinical utility of actionable molecular diagnostic tests in adult oncology. *Cent Med Technol Policy* [http://www.cmpnet.org/docs/resources/MDX\\_EGD.pdf](http://www.cmpnet.org/docs/resources/MDX_EGD.pdf). Accessed 28 July 2015.
5. 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*. Oxford University Press: New York, 2003:217–233.
6. Moore HM, Kelly AB, Jewell SD, et al. Biospecimen reporting for improved study quality (BRISQ). *Cancer Cytopathol* 2011;119:92–101.
7. Altman DG, McShane LM, Sauerbrei W, Taube SE. Reporting recommendations for tumor marker prognostic studies (REMARK): explanation and elaboration. *BMC Med* 2012;10:51.
8. Sun F, Bruening W, Erinoff E, Schoelles KM. *Addressing Challenges in Genetic Test Evaluation: Evaluation Frameworks and Assessment of Analytic Validity*. Agency for Healthcare Research and Quality (AHRQ): Rockville, MD, 2011.
9. McShane LM, Hayes DF. Publication of tumor marker research results: the necessity for complete and transparent reporting. *J Clin Oncol* 2012;30:4223–4232.
10. Freidlin B, McShane LM, Korn EL. Randomized clinical trials with biomarkers: design issues. *J Natl Cancer Inst* 2010;102:152–160.
11. Simon R. Clinical trial designs for evaluating the medical utility of prognostic and predictive biomarkers in oncology. *Per Med* 2010;7:33–47.
12. Mandrekar SJ, Sargent DJ. Clinical trial designs for predictive biomarker validation: one size does not fit all. *J Biopharm Stat* 2009;19:530–542.
13. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. *BMJ* 2009;338:b604.
14. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009;338:b605.
15. Janes H, Pepe MS, Bossuyt PM, Barlow WE. Measuring the performance of markers for guiding treatment decisions. *Ann Intern Med* 2011;154:253–259.
16. Simon R. Development and validation of therapeutically relevant multi-gene biomarker classifiers. *J Natl Cancer Inst* 2005;97:866–867.
17. Pepe MS, Janes HE. Gauging the performance of SNPs, biomarkers, and clinical factors for predicting risk of breast cancer. *J Natl Cancer Inst* 2008;100:978–979.
18. Pepe MS, Janes H, Gu JW. Letter by Pepe et al regarding article, “Use and misuse of the receiver operating characteristic curve in risk prediction”. *Circulation* 2007;116:e132; author reply e134.
19. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ* 2009;338:b606.
20. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ* 2009;338:b375.
21. Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003;3:25.
22. Institute of Medicine. *Evolution of Translational Omics: Lessons Learned and the Path Forward*. National Academies Press: Washington, DC, 2012.
23. Lord SJ, Irwig L, Bossuyt PM. Using the principles of randomized controlled trial design to guide test evaluation. *Med Decis Making* 2009;29:E1–E12.
24. Basch E, Abernethy AP, Mullins CD, et al. Recommendations for incorporating patient-reported outcomes into clinical comparative effectiveness research in adult oncology. *J Clin Oncol* 2012;30:4249–4255.
25. US Food and Drug Administration. *Guidance for industry, enrichment strategies for clinical trials to support approval of human drugs and biological products*. 2012. <http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm332181.pdf>. Accessed 28 June 2015.
26. Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J Natl Cancer Inst* 2009;101:1446–1452.
27. Berger ML, Dreyer N, Anderson F, Towse A, Sedrakyan A, Normand SL. Prospective observational studies to assess comparative effectiveness: the ISPOR good research practices task force report. *Value Health* 2012;15:217–230.
28. 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.
29. Glicklich R, Dreyer N, Leavy N, eds. *Registries for Evaluating Patient Outcomes: A User's Guide*. 3rd edn. Agency for Healthcare Research and Quality: Rockville, MD, 2014.
30. Velentgas P, Dreyer N, Nourjah P, eds. *Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide*. AHRQ publication no. 12(13)-EHC099. Agency for Healthcare Research and Quality: Rockville, MD, 2013.
31. Dreyer NA, Schneeweiss S, McNeil BJ, et al.; GRACE Initiative. GRACE principles: recognizing high-quality observational studies of comparative effectiveness. *Am J Manag Care* 2010;16:467–471.
32. Roberts M, Russell LB, Paltiel AD, Chambers M, McEwan P, Krahn M; ISPOR-SMDM Modeling Good Research Practices Task Force. Conceptualizing a model: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-2. *Med Decis Making* 2012;32:678–689.
33. 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;27:76–82.
34. Blue Cross Blue Shield Technology Evaluation Center. November 2011. [http://www.bcbs.com/blueresources/tec/vols/26/26\\_07.pdf](http://www.bcbs.com/blueresources/tec/vols/26/26_07.pdf).
35. Parkinson DR, McCormack RT, Keating SM, et al. Evidence of clinical utility: an unmet need in molecular diagnostics for patients with cancer. *Clin Cancer Res* 2014;20:1428–1444.
36. McShane LM, Cavenagh MM, Lively TG, et al. Criteria for the use of omics-based predictors in clinical trials: explanation and elaboration. *BMC Med* 2013;11:220.
37. Febbo P, Ladany M, Aldape K, 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–S32; quiz S 33.
38. Lyman GH, Levine M. Comparative effectiveness research in oncology: an overview. *J Clin Oncol* 2012;30:4181–4184.
39. Center for Medical Technology Policy. *Initial Medical Policy and Model Coverage Guidelines for Clinical Next Generation Sequencing in Oncology, Report and Recommendations*. August 2015. [http://www.cmpnet.org/docs/resources/Full\\_Release\\_Version\\_August\\_13\\_\\_2015.pdf](http://www.cmpnet.org/docs/resources/Full_Release_Version_August_13__2015.pdf). Accessed 28 August 2015.
40. Institute of Medicine. *Generating Evidence for Genomic Diagnostic Test Development: Workshop Summary*. National Academies Press: Washington, DC, 2011.