

## Building the evidence base for decision making in cancer genomic medicine using comparative effectiveness research

Katrina A.B. Goddard PhD<sup>1</sup>, William A. Knaus MD<sup>2</sup>, Evelyn Whitlock MD<sup>1</sup>, Gary H. Lyman MD, MPH<sup>3</sup>, Heather Spencer Feigelson PhD<sup>4</sup>, Sheri D. Schully PhD<sup>5</sup>, Scott Ramsey MD, PhD<sup>6</sup>, Sean Tunis MD, MSc<sup>7</sup>, Andrew N. Freedman PhD<sup>5</sup>, Muin J. Khoury MD PhD<sup>5,8</sup> and David L. Veenstra PharmD, PhD<sup>9</sup>

The clinical utility is uncertain for many cancer genomic applications. Comparative effectiveness research (CER) can provide evidence to clarify this uncertainty. The aim of this study was to identify approaches to help stakeholders make evidence-based decisions and to describe potential challenges and opportunities in using CER to produce evidence-based guidance. We identified general CER approaches for genomic applications through literature review, the authors' experiences, and lessons learned from a recent, seven-site CER initiative in cancer genomic medicine. Case studies illustrate the use of CER approaches. Evidence generation and synthesis approaches used in CER include comparative observational and randomized trials, patient-reported outcomes, decision modeling, and economic analysis. Significant challenges to conducting CER

in cancer genomics include the rapid pace of innovation, lack of regulation, and variable definitions and evidence thresholds for clinical and personal utility. Opportunities to capitalize on CER methods in cancer genomics include improvements in the conduct of evidence synthesis, stakeholder engagement, increasing the number of comparative studies, and developing approaches to inform clinical guidelines and research prioritization. CER offers a variety of methodological approaches that can address stakeholders' needs and help ensure an effective translation of genomic discoveries.

*Genet Med* 2012;14(7):633–642

**Key Words:** clinical utility; evidence generation; evidence synthesis; stakeholder

Clinical validity—the association between genotype and clinical phenotype—is now available for an increasing number of genomic applications. On the other hand, clinical utility—the improvement in patient outcomes and balance of risks and benefits—is largely unknown for most genomic applications. Implementing tests with uncertain clinical utility potentially wastes health-care resources through variable or unnecessary use of those tests. In the worst case, individuals are harmed when they or their health-care provider acts on the test results and they receive ineffective, potentially harmful treatments, or the results cause anxiety or discrimination. Furthermore, clinical utility may be quite specific, as when limited to subgroups with certain genotypes.<sup>1</sup> To maximize the clinical relevance of existing and as-yet-unknown genomic applications, it is crucial to ensure that clinically valid tests also have high clinical utility before they become widely used.

Clinical utility may be unclear for numerous reasons, including the relative lack of regulatory requirements for test manufacturers.<sup>2</sup> Furthermore, the research community has not aggressively prioritized either the translation of new discoveries into practical use or the generation of evidence with respect to these applications.<sup>3</sup> The field is also changing so quickly that evidence becomes rapidly outdated. In some cases, there may

be little incentive for private-sector investment in molecular diagnostics because of a lack of value-based reimbursement. Finally, existing paradigms for generating and evaluating evidence may be too slow, too costly, too unwieldy, or too unrepresentative to provide useful evidence to decision makers in a timely manner.<sup>4–7</sup>

Comparative effectiveness research (CER) is intended to create evidence for decision making, and to find out “what works” in health care. Although many definitions of CER have been proposed,<sup>8–12</sup> we use the Institute of Medicine's definition:<sup>10</sup> “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.” Some also use the term “patient-centered outcomes research” to refer to this type of research, although this concept will ultimately carry its own definition.

Concerns over the growing costs of health care<sup>13–15</sup> have made the use of CER a practical necessity, which has been enabled by \$1.1 billion in funding from the American Recovery and Reinvestment Act, and the advent of the Patient-Centered

<sup>1</sup>Center for Health Research, Kaiser Permanente Northwest, Portland, Oregon, USA; <sup>2</sup>Department of Public Health Sciences, University of Virginia, Charlottesville, Virginia, USA; <sup>3</sup>Comparative Effectiveness and Outcomes Research, Duke University and the Duke Cancer Institute, Durham, North Carolina, USA; <sup>4</sup>Institute for Health Research, Kaiser Permanente Colorado, Denver, Colorado, USA; <sup>5</sup>Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, Maryland, USA; <sup>6</sup>Cancer Prevention Research Program, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA; <sup>7</sup>Center for Medical Technology Policy, Baltimore, Maryland, USA; <sup>8</sup>Office of Public Health Genomics, Centers for Disease Control and Prevention, Atlanta, Georgia, USA; <sup>9</sup>Pharmaceutical Outcomes Research and Policy Program, Department of Pharmacy, University of Washington, Seattle, Washington, USA. Correspondence: Katrina Goddard ([katrina.ab.goddard@kpchr.org](mailto:katrina.ab.goddard@kpchr.org))

Submitted 4 October 2011; accepted 19 January 2012; advance online publication 17 April 2012. doi:10.1038/gim.2012.16

Outcomes Research Institute (<http://pcori.org>) in the 2010 Patient Protection and Affordable Care Act. Other developments that make CER timely are a new genetic test registry (<http://www.ncbi.nlm.nih.gov/gtr/>) at the National Institutes of Health, recent congressional hearings stimulated by concerns over direct-to-consumer genetic testing in July 2010, and possible changes at the Food and Drug Administration to consider genetic tests as medical devices, which would require regulatory approval before marketing.

It is critical that all stakeholders (including consumers, insurers, policymakers, and clinicians) possess tools to assess the clinical utility of genomic applications. We describe CER approaches to answer questions about cancer genomic applications, and the potential challenges and opportunities associated with each. We provide case studies of genomic applications to illustrate the types of questions decision makers are facing, and describe potential CER study designs and methods that can be used to address them.

**METHODS**

We searched PubMed for recent literature on CER and searched the citations of these articles to identify additional publications relevant to CER. We also considered additional articles that were not identified through this search but were known to the authors. We selected the following methodology categories for consideration: evidence synthesis, prospective comparative clinical trials, observational research, health economics and decision modeling, and stakeholder engagement. We developed descriptions of these approaches as applied to CER based on literature reviews and the authors' experience. We then identified a series of case studies of breast cancer genomic applications to clarify CER questions and possible methods to address them. We selected breast cancer because of the public health relevance of the disease, and because of the plethora of genomic applications currently in clinical practice. We used the ACCE framework (analytic validity, clinical validity, clinical utility, and ethical, legal, and social implications)<sup>16</sup> as a starting point to identify and organize the information we would abstract on the case studies. Finally, we identified particular challenges

for using these CER approaches to conduct genome-based research.

**RESULTS**

Our results are presented in three sections: (i) identification of the key questions for CER applications in cancer genomics, (ii) illustration of the key questions using examples from breast cancer genomic applications, and (iii) general methodological approaches to addressing the key questions.

**Key questions**

We framed our analysis using key questions in four areas, which are drawn from the ACCE framework<sup>16</sup> and other models.<sup>17</sup>

- Is there a significant association between the results of the genomic application and clinical phenotype? (clinical validity)
- Does the genomic application provide correct information? (analytic validity)
- Does the genomic application provide clinically significant information? (clinical utility)
- Does the genomic application lead to improved patient outcomes as compared with the alternative? (comparison or added clinical value)

**Illustration of key questions using cancer genomic applications as examples**

Genomic applications can span the entire range of disease, from risk identification to diagnosis and patient management. **Table 1** shows examples of both conventional and genomic applications in the context of breast cancer for each test category. We provide summary tables of example key questions for breast cancer genomic applications that address risk assessment (**Table 2**) and treatment decisions (**Table 3**).

*Clinical validity* is the association between the predictor (e.g., genotype, profile, or family history status) and clinical phenotype. Predictors are identified by investigating targeted pathways, by candidate-gene analysis, or through agnostic genome-wide study designs. Methodological problems from multiple testing, heterogeneity, the “winner’s curse” (the likelihood that the first report of a significant test will have a larger

**Table 1 Test categories and relationship to breast cancer disease status**

Disease status	Test category	Genomic test example	Conventional example
Predisease	Predictive risk assessment	<i>BRCA1/2</i> Family history SNP genotypes	Age, behavioral factors (e.g., nulliparity)
Asymptomatic disease	Screening	None	Mammography
Disease diagnosis	Differential diagnosis	None	Tumor histopathology
Disease management and treatment	Prognosis	<i>HER2</i> genotype Gene expression profile	Estrogen, progesterone receptor
	Pharmacogenomic	<i>HER2</i> genotype <i>CYP2D6</i> genotype	Estrogen, progesterone receptor
	Monitoring	None	Tumor markers (cancer antigen CA 15-3, CA 27.29; carcinoembryonic antigen)

SNP, single-nucleotide polymorphism.

effect size than later replication studies), small sample size, and other concerns make interpretation challenging.<sup>18–20</sup> Further, the attributable risk may be small because of low frequency or low penetrance, or the measured variant may only be linked to the functional variant. For example, initial studies reported an association between *CYP2D6* variants and the risk of disease recurrence in women taking tamoxifen (Table 3).<sup>21</sup> A systematic evidence review, however, found inconsistent evidence.<sup>22</sup> Preliminary results from recent retrospective analyses of large randomized controlled trials (RCTs) including about 5,000 women<sup>23,24</sup> found no association between *CYP2D6* variants and breast cancer recurrence.

*Analytic validity* refers to characteristics of the test, including reproducibility (i.e., will the same test performed on the same sample produce the same result?), the lower limit of detection (smallest quantity of the target that can be reliably detected), and analytic specificity (ability to measure the target and only the target). A proficiency testing program (exchange of quality control material for analysis and comparison across laboratories) may be the best approach to address this concern. For example, when *HER2* testing (Table 3) was first used in breast cancer clinical trials, it is estimated that up to 20% of test results may have been incorrect. Laboratories with lower volume testing were the most likely to report incorrect findings.<sup>25,26</sup>

**Table 2 Risk assessment genomic applications: summary of current evidence for breast cancer case studies**

Test application	Analytic validity	Clinical validity	Comparative effectiveness research	
			Clinical utility	Added clinical value
Family history (FHx)	<p>QUESTION: How well can we measure (test application)?</p> <ul style="list-style-type: none"> <li>RESULT: Currently not well documented or collected in either primary care or specialty practice. Only a minority of abstracted charts find a documented FHx.</li> <li>APPROACH: Observational studies; chart review</li> <li>LIMITATIONS: Small family size, predominance of males, adoption, lack of information about relatives, misattributed paternity, skewed segregation, and incomplete penetrance limit utility of the tool.</li> </ul>	<p>QUESTION: Is (test application) associated with the risk of developing breast cancer (BC)?</p> <ul style="list-style-type: none"> <li>RESULT: There is a strong but not universal demonstrated association between FHx and development of BC. A substantial proportion of confirmed <i>BRCA</i> mutation carriers lack a FHx in first- and second-degree relatives.</li> <li>APPROACH: Association lineage studies; observational studies.</li> <li>LIMITATIONS: Only about 20% of US women who develop breast cancer have a positive FHx. FHx is not always defined the same way.</li> </ul>	<p>QUESTION: Are there actions based on (test application) to prevent BC? What are the benefits and harms?</p> <ul style="list-style-type: none"> <li>RESULT: Risk reduction surgery (prophylactic mastectomy and oophorectomy) and chemoprevention for patients at high risk because of FHx. Women with “only” positive FHx have had these prophylactic procedures.</li> <li>APPROACH: Observational case–control or historical control studies.</li> <li>LIMITATIONS: Limited evidence of the positive impact of these actions on patient’s quality of life or other important outcomes.</li> </ul>	<p>QUESTION: Is use of (test application) better than the alternative for predicting BC risk?</p> <ul style="list-style-type: none"> <li>COMPARATOR: Risk assessment tools using clinical characteristics (e.g., breast density) or SNP panels.</li> <li>RESULT: FHx is consistently the strongest predictor of BC risk. None of the “expanded” risk models have ever excluded FHx from the final models.</li> <li>APPROACH: Observational studies usually using r2 or sensitivity and specificity as measures of discrimination.</li> <li>LIMITATIONS: No widespread comparisons made with GWAS.</li> </ul>
SNP genotypes	<ul style="list-style-type: none"> <li>RESULT: Analytic validity is high. Genotypes can be reliably measured through array-based genotyping in GWAS.</li> <li>APPROACH: Array-based genotyping.</li> <li>LIMITATIONS: Common variants have a much higher coverage than rarer ones; error rates are higher for SNPs with a low minor allele frequency.</li> </ul>	<ul style="list-style-type: none"> <li>RESULT: 18 SNPs are significantly associated with BC and have been replicated in separate studies.</li> <li>APPROACH: GWAS</li> <li>LIMITATIONS: Although some of the initial prediction results using retrospective analyses appear promising, the work is not able to answer fundamental questions such as whether a given set of SNPs can explain a large proportion of familial risk. No prospective cohort study.</li> </ul>	<ul style="list-style-type: none"> <li>RESULT: Unknown</li> <li>APPROACH: None</li> <li>LIMITATIONS: No clinical tool exists and therefore no studies exist on the clinical utility of these tests in better managing BC risk in patients, nor of unintended consequences.</li> </ul>	<ul style="list-style-type: none"> <li>COMPARATOR: Risk assessment without SNP genotypes</li> <li>RESULT: SNP genotypes modestly improve the performance of risk models.</li> <li>APPROACH: Area under the receiver operator characteristics curve</li> <li>LIMITATIONS: The identified SNPs have small and unvalidated incremental value as compared with existing models.</li> </ul>

GWAS, genome-wide association studies; r2, coefficient of determination statistic; SNP, single-nucleotide polymorphism.

**Table 2 Continued on next page.**

Table 2 Continued.

	Analytic validity	Clinical validity	Comparative effectiveness research	
			Clinical utility	Added clinical value
<i>BRCA1/2</i> genotype	<ul style="list-style-type: none"> <li>• RESULT: Analytic validity is high. Myriad genetics reported analytic sensitivity and analytic specificity &gt;99%.</li> <li>• APPROACH: Observational studies</li> <li>• LIMITATIONS: Some occult mutations are not detected by current methods, and there undoubtedly remain undiscovered mutations.</li> </ul>	<ul style="list-style-type: none"> <li>• RESULT: Clinical validity is high for deleterious mutations, and negative results for known familial mutations. For the first family member to be tested, a “no mutation identified” result is indeterminate because not all <i>BRCA</i> mutations can be identified; 13% of tests result in variants of “uncertain clinical significance.”</li> <li>• APPROACH: Observational studies</li> <li>• LIMITATIONS: Women who carry a variant of uncertain clinical significance are managed according to the personal and family history.</li> </ul>	<ul style="list-style-type: none"> <li>• RESULT: Intensive screening of women with familial risk shows improved detection but considerable interval cancers still occur.</li> <li>• Chemoprevention with tamoxifen has benefit for high-risk women, and specifically for <i>BRCA</i> carriers.</li> <li>• Mastectomy conferred BC risk reduction ranging from 85–100%. Oophorectomy conferred ovarian cancer risk reduction ranging from 85–100%, and 53–68% for BC when done before menopause.</li> <li>• High-risk women may use more effective BC screening modality.</li> <li>• APPROACH: Observational studies</li> <li>• LIMITATIONS: Few screening or prevention recommendations for ovarian cancer among women with <i>BRCA</i> mutations exist.</li> </ul>	<ul style="list-style-type: none"> <li>• COMPARATOR: Risk assessment tools without <i>BRCA1/2</i> genotype</li> <li>• RESULT: <i>BRCA1/2</i> genotyping is advantageous to identify high-risk women in some categories identified by personal and family history of breast or ovarian cancer, age at diagnosis, and Ashkenazi Jewish ancestry. Testing is unlikely to be cost effective in categories of women with low <i>a priori</i> risk.</li> <li>• APPROACH: Observational studies, modeling</li> <li>• LIMITATIONS: The majority of BC cases are not explained by <i>BRCA1/2</i> mutations.</li> </ul>

GWAS, genome-wide association studies; SNP, single-nucleotide polymorphism.

A proficiency testing program has since been implemented for *HER2*.<sup>27</sup>

*Clinical utility* has to do with whether the information provided by the genomic application is actionable, and with evaluating the balance between risks and benefits of available actions. *BRCA1/2* testing (Table 2) is one example. Mutation carriers are at increased risk of developing breast and ovarian cancer and can receive more effective breast cancer screening by choice of screening modality or interval, can undergo surgeries to reduce risk by 85–100%, or can select chemoprevention. Women at high risk in families with known mutations who undergo testing and are found not to carry deleterious *BRCA1/2* mutations can receive significant psychosocial benefit and avoid these interventions. On the other hand, the clinical utility of gene expression profiles is less clear.<sup>28</sup> A key area of uncertainty is how women and their physicians will make treatment decisions based on test results in the intermediate risk category. Two prospective RCTs—TAILORx and RxPONDER—are under way to evaluate how risk profile scores affect patient management, treatment decisions, and subsequent outcomes.<sup>29,30</sup>

*Added clinical value*<sup>17</sup> asks whether the application provides better clinical, patient, or economic outcomes than those of the alternative, which could be another intervention or usual care.

A critical factor is how to define and measure “better,” which could include measures of predictive accuracy, quality of life, survival, or other outcomes, including testing costs, acceptability, or feasibility. Recently, a genetic risk prediction model for breast cancer was published including 10 well-validated single-nucleotide polymorphisms (Table 2).<sup>31</sup> The predictive power of this genetic model is only slightly better (about 4%) than the widely used Gail model,<sup>32</sup> which uses nongenetic factors to predict risk. Because both models explain about 60% of risk, and because the Gail model can be used without the expense of genetic testing, the added clinical value of the risk prediction model based on single-nucleotide polymorphism profiles is low.

**General methods for comparative effectiveness research**

The key questions and methodological challenges described earlier, coupled with the need for CER to inform a diverse group of stakeholders, will require a range of innovative strategies, including both evidence synthesis and evidence generation (Table 4).

**Synthesis of existing evidence**

Evidence synthesis begins with identifying topics through processes such as horizon scanning,<sup>33</sup> which searches published literature and gray literature databases (e.g., meeting abstracts,

**Table 3 Pharmacogenomic applications: summary of current evidence for breast cancer case studies**

Test application	Analytic validity	Clinical validity	Comparative effectiveness research	
			Clinical utility	Added clinical value
<i>HER2</i> genotype and trastuzumab treatment response	<ul style="list-style-type: none"> <li>QUESTION: How well can we measure (test application)?</li> <li>RESULT: Test characteristics depend on whether IHC or FISH platform is used, test volume of the laboratory, and testing strategy (e.g., FISH or IHC test first).</li> <li>APPROACH: Observational studies</li> <li>LIMITATIONS: <i>HER2</i> status may change when comparing the primary tumor to metastases.</li> </ul>	<ul style="list-style-type: none"> <li>QUESTION: Is (test application) associated with the treatment response?</li> <li>RESULT: <i>HER2</i> status is associated with survival following treatment with trastuzumab in early-stage and metastatic disease, and with response to anthracyclines.</li> <li>APPROACH: Randomized clinical trials</li> <li>LIMITATIONS: Some studies on anthracyclines had small sample sizes</li> </ul>	<ul style="list-style-type: none"> <li>QUESTION: What are the benefits and harms of (test application) for treatment decisions?</li> <li>HARMS: Potential for trastuzumab to be effective for some (small %) patients who are <i>HER2</i> negative</li> <li>BENEFITS: Avoiding cardiotoxic side effects in patients who will not respond to trastuzumab.</li> <li>Avoiding use of an expensive treatment for patients who will not benefit from treatment</li> </ul>	<ul style="list-style-type: none"> <li>QUESTION: Is use of (test application) better than the alternative?</li> <li>COMPARATOR: Herceptin treatment decisions in the absence of knowledge of <i>HER2</i> status</li> <li>RESULT: None.</li> <li>APPROACH: None.</li> <li>LIMITATIONS: None.</li> </ul>
<i>CYP2D6</i> genotype and tamoxifen treatment response	<ul style="list-style-type: none"> <li>RESULT: High for SNP detection</li> <li>APPROACH: Observational studies</li> <li>LIMITATIONS: None.</li> </ul>	<ul style="list-style-type: none"> <li>RESULT: Not well established; conflicting results</li> <li>APPROACH: Retrospective analysis of prospective clinical trials and retrospective cohort studies</li> <li>LIMITATIONS: Small to medium size studies; conflicting</li> </ul>	<ul style="list-style-type: none"> <li>HARMS: Unknown.</li> <li>BENEFITS: Alternative treatments for women with poor metabolizer genotypes (aromatase inhibitors)</li> <li>Tamoxifen could become the preferred effective and inexpensive therapy for many women without poor metabolizer genotypes</li> </ul>	<ul style="list-style-type: none"> <li>COMPARATOR: Treatment decisions in the absence of knowledge of <i>CYP2D6</i> genotype</li> <li>RESULT: None.</li> <li>APPROACH: None.</li> <li>LIMITATIONS: Direct assessment of clinical utility with RCTs expensive</li> </ul>

FISH, fluorescence in situ hybridization; IHC, immunohistochemistry; RCT, randomized controlled trial; SNP, single-nucleotide polymorphism.

commercial websites, newsletters, or business news) for emerging genomic applications. Horizon scanning may also examine existing curated databases of published literature such as the HuGE Navigator, the GAPP Knowledge Base, and the Pharmacogenomics Knowledge Base. Gray literature sources identify emerging genomic applications because of the lag in reporting on these topics in peer-reviewed published literature; these may be supplemented by a query process from users as an early indicator of burgeoning clinical interest. Once new topics are identified, rapid topic briefs, or short reviews, are used to assess the feasibility of a full systematic review.

Full systematic reviews are often identified through a public nomination process and then commissioned through an existing body such as the U.S. Preventive Services Task Force, Evaluation of Genomic Applications in Practice and Prevention Working Group, or the Agency for Healthcare Research and Quality Effective Healthcare Program. The scope of the review is defined by the analytic framework and key questions, and the reviewers conduct a broad but systematic search to identify evidence. They develop inclusion and validity criteria for the evidence, and abstract needed data, which is then synthesized and summarized in a narrative. Quantitative approaches such as meta-analysis may provide summary estimates of critical measures across studies. Although full systematic reviews are

comprehensive, they may not be timely, which is a critical issue in summarizing evidence in genomics.

**Generation of new evidence**

*Clinical trials.* Explanatory RCTs are used to evaluate the efficacy of a medical intervention. They are often viewed as the ideal approach to protect against bias. However, this study design also has limitations.<sup>34,35</sup> Explanatory RCTs are typically restricted to selected patients, but real-world populations can differ markedly in age, race, comorbid conditions, concomitant medication use, and environmental factors. The generally small sample size of RCTs may underrepresent some patient groups, a particular concern when evaluating genomic-based subgroups. Randomization requires a prospective design, and so RCTs tend to focus on questions of short-term efficacy and safety using intermediate (surrogate) end points. Finally, because RCT protocols are often far removed from routine practice, they may not accurately predict real-world effectiveness.

Innovative strategies in the design of clinical trials seek to overcome these limitations. Pragmatic clinical trials<sup>36,37</sup> address the issue of relevance by assessing the effectiveness of the intervention in routine practice by using wide patient inclusion criteria, allowing variation in the treatment protocol, and assessing outcomes relevant to everyday life. However, these studies typically



**Table 4 Opportunities in comparative effectiveness research**

Type of study	Methods and opportunities
Evidence synthesis	Horizon scanning Rapid/state-of-the-science reviews Identification and prioritization of research gaps Systematic reviews Health technology assessment
Randomized controlled trials (RCTs)	Explanatory RCTs Bayesian and adaptive clinical trials Cluster randomized trials Pragmatic clinical trials Study of subpopulations/subgroup analyses Incorporation of genetic questions/information into trial design Proxy outcome measures Coverage with evidence development
Observational studies	Case-control design Retrospective and prospective cohort design Quasi-experimental design Administrative databases and electronic medical records Retrospective analysis of biospecimens from RCTs Selection bias: instrumental variables, propensity scores Confounding by indication: restriction
Decision modeling/economic analysis	Value of information analysis/scenario modeling Risk-benefit modeling to facilitate evaluation of indirect evidence Stakeholder engagement and knowledge brokering (GAPPNet)

GAPPNet, Genomic Applications in Practice & Prevention Network.

require much larger sample sizes to compensate for heterogeneity in the patient population and the treatment protocol, and longer time frames to assess patient-relevant outcomes.

To fund and implement studies with larger sample sizes, collaborations between researchers, health-care systems, and payers will be critical. A policy framework for conducting such collaborations is coverage with evidence development. Coverage with evidence development is a conditional reimbursement decision by a payer, with an explicit linkage between payment and data collection to reduce uncertainty about the intervention.<sup>38,39</sup> The Centers for Medicare and Medicaid Services recently issued a coverage with evidence development policy for warfarin pharmacogenomic testing, in which the Centers will pay for testing if the patient is enrolled in a RCT designed to measure bleeding events.<sup>40</sup>

Cluster randomized trials are another alternative experimental design in which units such as communities, medical clinics or hospitals, or families are randomized to intervention arms rather than individuals. This design is often used when the intervention is aimed at changing the behavior of the group or the behavior of a provider, or changing the organization of services. This design can also be used to reduce contamination (e.g., “spillover” effects of a mass educational campaign), or

to improve the feasibility of a study. Cluster randomized trials require more sophisticated analytic approaches and larger sample sizes because of lack of independence among individual observations.<sup>41,42</sup> However, this study design may still be cost-efficient.<sup>43</sup> Cluster randomized trials have been used to assess the impact of decision support tools implemented at the provider level, particularly involving genetic risk assessment based on family history.<sup>44-46</sup>

Bayesian or adaptive trial designs can accelerate the pace of evidence generation by incorporating information from prior cases to alter the study midway, based on interim results. An adaptive design incorporates genomic profiles into the trial design by changing the patient randomization process to treatment arms as the trial progresses based on the accumulated data for each profile.<sup>47</sup> Despite potential advantages, these trials have not gained widespread acceptance because of nonstandard methods and resistance among Food and Drug Administration regulators.

One example of an adaptive design is the I-SPY 2 project.<sup>48</sup> This is a phase II RCT in the neoadjuvant setting for women with locally advanced breast cancer. Patients are randomized to treatment arms based on their biomarker profile. Initially, patients with a given biomarker profile have an equal chance of being randomized to each treatment arm. Over time, the randomization ratio (i.e., the vector of probabilities that a patient will be randomized to each treatment arm) for each biomarker profile is adjusted depending on the experience of previously randomized patients with that profile. Thus, future patients are more likely to be randomized to treatment arms in which patients with similar biomarker profiles achieved a better response.

*Observational studies.* Observational study designs are a valuable and complementary approach to RCTs.<sup>34,35,49</sup> These designs are especially useful when it would be unethical or infeasible to conduct an RCT. For example, Habel and colleagues (2006)<sup>50</sup> conducted a retrospective case-control study to evaluate the association between long-term outcomes (the risk of breast cancer death) and Oncotype DX Recurrence Score. Previous studies based on RCTs could not evaluate this outcome and used shorter-term outcomes instead, including rates of distant recurrence as the primary measures.<sup>51,52</sup> The primary limitation with observational study designs is the possibility of confounding bias due to unexplained differences between exposure groups, which are not controlled for through randomization. One option is to use risk-adjustment approaches, such as propensity scores or instrumental variables. However, unlike randomization, these approaches cannot control for unmeasured or imperfectly measured covariates, so residual confounding may still be present. Observational designs are less subject to bias when there is no relationship between treatment assignment and treatment response, and they can contribute important information about unanticipated, real-world impacts that complements RCTs.

The use of large, administrative health-care databases to access routinely collected data may offer significant advantages

for an observational design. The large population size enables the study of infrequent events. Also, such databases are representative of routine care, making it possible to study real-world effectiveness and utilization patterns. The data are available at relatively low cost without long delays as compared with data-gathering for a new prospectively recruited study. Electronic data from integrated health-care systems with a defined population and electronic medical records (EMRs) allow broad consideration of the patient's health status. Over time, EMRs and associated databases will make it feasible to consider long-term outcomes. Challenges with the use of EMRs for research include (i) much of the data is in unstructured notes, requiring manual abstraction or natural language processing, (ii) there is a lack of harmonization across systems because of multiple or lacking data standards, (iii) there may be discontinuity of longitudinal data for patients depending on the source or access to health insurance, and (iv) there is variable data quality because of the multitude of providers who enter data into the system. A specific limitation is a lack of clinically derived genomic information or the ability to easily access it.<sup>53</sup> Although these challenges may limit use of data for research from the majority of health-care providers in the United States with EMRs at the moment, nevertheless, there are examples of systems that are currently able to use EMR data for research and that have biorepositories linked to EMRs to facilitate retrospective study designs.

*Decision modeling and health economics.* Evidence-based bodies have generally relied on RCTs to inform their guideline development when weighing relative benefits and harms. Decision modeling provides a framework to formally incorporate indirect and direct evidence from various sources, to evaluate likely outcomes, and to quantify uncertainty. The advantages of this approach are a structured, transparent framework for assessing the available evidence, and, critically, for quantifying the uncertainty of evidence and its potential impact on patient outcomes. Challenges include timeliness of implementation, development of models acceptable to stakeholders, problems with assumptions and model transparency, and the development of formal guidelines or recommendations based on modeling analyses. Recent work indicates that stakeholders such as clinicians, health-care payers, and guidelines groups are open to using such approaches in genomics if the process is transparent and there is not an overreliance on the model results to drive recommendations.<sup>54</sup>

Another CER approach is value-of-research analysis, also called value of information analysis, which is used to make decisions about selecting technologies for additional research trials and for designing those trials optimally. The concept behind value of research is that additional research reduces our uncertainty about which intervention to use in clinical practice.<sup>55</sup> Reducing uncertainty is valuable because it reduces the chances that the less optimal strategy is selected, and studies that provide "negative" results are still valuable. Impacts on patients' morbidity and mortality are assessed, as well as health-care costs.

These approaches are just beginning to be applied to research prioritization decisions in health care, and must be shown to be feasible as well as useful before widespread implementation. The value of research paradigm may be particularly useful in genomics because the pace of innovation leads to the need to prioritize investment in expensive comparative studies.<sup>56</sup>

Cost-effectiveness analysis is the standard approach to formally assessing the incremental value of health-care technologies.<sup>57</sup> These analyses can incorporate a variety of outcomes including clinical events, life expectancy, quality-adjusted life expectancy, and health-care costs. Applying cost-effectiveness analysis to genomics can be challenging. First, the general lack of comparative effectiveness data makes evaluation of comparative value problematic, and uncertainty must be carefully assessed. Second, the value patients and clinicians place on knowing genetic information (the "value of knowing") is difficult to measure and to incorporate into policy decisions.<sup>58,59</sup> Contingent valuation (willingness-to-pay) approaches have been used;<sup>60</sup> more recently, discrete-choice experiments to assess patient preferences have offered significant promise.<sup>61</sup>

*Stakeholder engagement.* Given CER's explicit purpose of producing useful information for decision making, there has been increasing recognition of the importance of including stakeholders such as patients, clinicians, payers, and policymakers in CER activities. The Institute of Medicine recommended specifically that this work "should fully involve consumers, patients, and caregivers in ... strategic planning, priority setting, research proposal development, peer review, and dissemination."<sup>10</sup> The rationale is that such involvement will lead to a focus on questions of most relevance to end users.<sup>62</sup> Stakeholder involvement should increase the chances that study designs will reflect the specific questions of decision makers, and the greater relevance of the research questions will also facilitate use of results in decision making. Recent work by Deverka and colleagues is one example of an approach to involve stakeholders in assessing the current state of evidence.

Although the need for stakeholder engagement is widely recognized, the published literature on this topic is limited, and there are few formal evaluations of these methods.<sup>63</sup> Some qualitative synthesis has identified several recurring themes, including the importance of developing trust and shared understanding through sustained interaction and devoting adequate time and resources to training and preparation.<sup>64</sup> The need for valid methods for engaging patients, consumers, and clinicians has been identified as a critical CER methods research priority.<sup>65</sup>

## DISCUSSION

The complexity of developing sufficient evidence for the clinical utility of cancer genomic applications offers opportunities for innovative applications of CER-based approaches. Diagnostic tests such as *BRCA1/2* genotyping or Oncotype DX generate information, so it is necessary to use study designs that take into account subsequent therapeutic decisions in determining the clinical impact of the test. Another challenge is to identify

and address all important subgroups. In the adaptive clinical trial design of the I-SPY2 project, the subgroups are identified ahead of time, but in other contexts it may be preferable to consider retrospective study designs if the subgroups are not known beforehand. The rapid pace of innovation in genomics means that studies must be extremely efficient and informed by stakeholder needs if the evidence is to remain timely and relevant. Potential solutions to the aforementioned problems include adaptive clinical trials, retrospective studies using EMRs, and decision-modeling approaches to assess indirect evidence. The variable definitions and paucity of data for clinical utility present another challenge. For example, the concept of personal utility, or the value of knowing the information, is clearly relevant for some decision makers and settings (e.g., direct-to-consumer marketing) but may not be relevant in a clinical context,<sup>60</sup> and the metrics for measuring personal utility are not well established.<sup>58,66</sup> However, stakeholder engagement and approaches to assessing patient preferences such as conjoint analysis may offer a way forward. In the following, we provide a summary of the implications for CER in cancer clinical genomics.

We believe a more comprehensive approach is needed to resolve questions about the clinical utility of genomic applications. Specifically, research is needed that considers more outcome measures, and that is conducted in settings that are relevant to more real-world clinical decisions than have been considered in the past. For example, [Table 2](#) highlights some of the limitations in our knowledge about clinical utility for existing applications in the context of breast cancer. A multitude of stakeholders should have a role in evidence generation. For example, health systems are needed to provide data and facilitate pragmatic trials, providers are needed to use genomic tests in the context of evidence generation, and test developers are needed to make tests available for collaborative study. Such an undertaking, however, will be resource intensive. Thus, a more comprehensive approach will provide clear priorities for CER to ensure that limited resources are used to resolve the most compelling questions. A more comprehensive approach would also engage stakeholders to ensure the study of pressing topics in real-world environments and should establish approaches for rapid evidence synthesis and quantitatively assess the value of prioritized research, considering the health and well-being of patients and the decision-making needs of other stakeholders.

Second, it may be necessary to establish an evidentiary framework to clearly define evidence standards, particularly for clinical utility. Existing frameworks in genomic medicine primarily build upon the ACCE framework<sup>16</sup> or the stages of translational medicine,<sup>67</sup> and there is no regulatory requirement that applies to all genetic tests. A primary limitation of existing frameworks is that they provide no standard threshold for what constitutes “necessary and sufficient” evidence. What is urgently needed now is to establish appropriate evidentiary thresholds for genomic test adoption; this will require a dialogue and interaction between evidence appraisers and end users to develop consensus. Furthermore, these thresholds need to include appropriate

study design criteria and recognize that an RCT is not desirable or feasible in every circumstance, and to establish when (not if) an observational study design and evidence of underlying biological mechanisms contribute to the evidentiary framework.<sup>68</sup> Beyond study designs, an evidentiary framework needs to cogently articulate the minimal evidence necessary before clinical application is warranted, taking into consideration the type of genomic application and its clinical context.

Third, strategies that are rapid, timely, and efficient are needed, given the fast pace of discovery in genomic-based approaches.<sup>69</sup> Innovative methods, such as evidence heuristics to classify genes and variants, that are capable of addressing whole-genome sequencing and decision modeling frameworks will help address this need.<sup>70,71</sup> New strategies will involve transformation of the research infrastructure to “learning systems” that allow continual addition to the evidence base. This approach will achieve greater efficiency through efforts such as establishing biorepositories or registries, linking EMR data or administrative databases to genomic information and creating quality-assured clinical data repositories, or improving standardized coding schemes for genomic applications.

Finally, any reforms of the evidentiary framework should uphold rigorous standards on the statistical validity of the research.<sup>72</sup> Although some study designs have a risk of greater uncertainty, we can make strategic choices about when such increased uncertainty is acceptable. We should improve the integrity and conduct of all study designs by using guidelines such as those provided in Strengthening the Reporting of Observational Studies in Epidemiology (STROBE), Consolidated Standards of Reporting Trials Statement (CONSORT), Strengthening the Reporting of Genetic Associations (STREGA), and Genetic Risk Prediction Studies (GRIPS). Also, we can describe how threats to validity are assessed in grading evidence, or require preregistry of the analysis plan for observational studies, as is currently done for RCTs, to reduce biases (including selective outcome reporting) or errors, such as those generated from multiple testing.

## Conclusion

Informed decision making in cancer clinical genomics through the development and application of comparative effectiveness research could accelerate the implementation of valuable genomic applications while avoiding harmful applications that can persist in clinical care, leading to waste or patient harm.

## ACKNOWLEDGMENTS

This study was supported in part by cooperative agreements funded by the American Recovery and Reinvestment Act from the National Cancer Institute including the CERGEN study (UC2 CA148471 to K.A.B.G., E.W., and H.S.F.), the CANCEGEN study (RC2 CA148570-01 to S.R. and D.L.V.), RC2CA148041-01 (to G.H.L.), Building a Genome Enabled Electronic Medical Record (UC2CA150911 to W.A.K.), and a cooperative agreement funded by the Centers for Disease Control and Prevention (5U18-GD000005-02 to D.L.V.).



## DISCLOSURE

D.L.V. reports serving as a consultant for Medco, Novartis Molecular Diagnostics, and Genentech, and is supported by the following genomics-related research grants: P50HG003374, RC2CA148570, and U01GM092676 from the National Institutes of Health and U18GD000005 from the Centers for Disease Control and Prevention.

## REFERENCES

- Limdi NA, Veenstra DL. Expectations, validity, and reality in pharmacogenetics. *J Clin Epidemiol* 2010;63:960–969.
- SACGHS. U.S. System of Oversight of Genetic Testing: A Response to the Charge of the Secretary of Health and Human Services. [http://oba.od.nih.gov/oba/SACGHS/reports/SACGHS\\_oversight\\_report.pdf](http://oba.od.nih.gov/oba/SACGHS/reports/SACGHS_oversight_report.pdf). Accessed 15 September 2011.
- Schully SD, Benedicto CB, Gillanders EM, Wang SS, Khoury MJ. Translational research in cancer genetics: the road less traveled. *Public Health Genomics* 2011;14:1–8.
- Lauer MS. The historical and moral imperatives of comparative effectiveness research. *Stat Med* 2010;29:1982–4; discussion 1996.
- Atkins D. Creating and synthesizing evidence with decision makers in mind: integrating evidence from clinical trials and other study designs. *Med Care* 2007;45(10 suppl 2):S16–S22.
- Gatsonis C. The promise and realities of comparative effectiveness research. *Stat Med* 2010;29:1977–1981; discussion 1996.
- Normand SL, McNeil BJ. What is evidence? *Stat Med* 2010;29:1985–1988; discussion 1996.
- Agency for Healthcare Research and Quality. Glossary of terms. <http://www.effectivehealthcare.ahrq.gov/index.cfm/glossary-of-terms/?pageaction=showterm&termid=118>. Accessed 23 January 2011.
- Congressional Budget Office. Research on the Comparative Effectiveness of Medical Treatments: Issues and Options for an Expanded Federal Role. <http://www.cbo.gov/ftpdocs/88xx/doc8891/12-18-ComparativeEffectiveness.pdf>. Accessed 22 January 2011.
- Institute of Medicine. Initial National Priorities For Comparative Effectiveness Research. <http://www.iom.edu/cerpriorities>. Accessed 14 September 2011.
- National Cancer Institute. Overview of Comparative Effectiveness Research (CER). <http://cancercontrol.cancer.gov/cer/overview.html>. Accessed 14 January 2011.
- U.S. Department of Health and Human Services. Draft Definition of Comparative Effectiveness Research for the Federal Coordinating Council. <http://www.hhs.gov/recovery/programs/cer/draftdefinition.html>. Accessed 4 February 2011.
- Institute of Medicine. Learning what works best: the nations need for evidence on comparative effectiveness in health care, 2007. <<http://www/iom.edu/ebm-effectiveness>>.
- Congress of the United States Congressional Budget Office. Research on the Comparative Effectiveness of Medical Treatments. Congress of the US Congressional Budget Office. Washington, DC, 2007.
- Truffer CJ, Keehan S, Smith S, et al. Health spending projections through 2019: the recession's impact continues. *Health Aff (Millwood)* 2010;29:522–529.
- Centers for Disease Control and Prevention. Genomic Testing: ACCE Model Process for Evaluating Genetic Tests. <http://www.cdc.gov/genomics/gtesting/ACCE/index.htm>. Accessed 20 January 2011.
- Goodman S, Dickerson K, Wilson R; CMTF. Effectiveness guidance document: gene expression profile tests for early stage breast cancer. [http://cmtfnet.org/documents/egd\\_ge.pdf](http://cmtfnet.org/documents/egd_ge.pdf). Accessed 25 July 2011.
- Ioannidis JP. Why most discovered true associations are inflated. *Epidemiology* 2008;19:640–648.
- Kraft P. Curses–winner's and otherwise—in genetic epidemiology. *Epidemiology* 2008;19:649–651; discussion 657.
- Chanock SJ, Manolio T, Boehnke M, et al.; NCI-NHGRI Working Group on Replication in Association Studies. Replicating genotype-phenotype associations. *Nature* 2007;447:655–660.
- Goetz MP, Rae JM, Suman VJ, et al. Pharmacogenetics of tamoxifen biotransformation is associated with clinical outcomes of efficacy and hot flashes. *J Clin Oncol* 2005;23:9312–9318.
- Terasawa T, Dahabreh I, Castaldi P, Trikalinos T. Systematic Reviews on Selected Pharmacogenetic Tests for Cancer Treatment: CYP2D6 for Tamoxifen in Breast Cancer, KRAS for anti-EGFR antibodies in Colorectal Cancer, and BCR-ABL1 for Tyrosine Kinase Inhibitors in Chronic Myeloid Leukemia. Technology Assessment Report Project ID: GEN0609. <http://www.cms.gov/DeterminationProcess/downloads/d76TA.pdf>. Accessed 16 September 2011.
- Rae JM, Drury S, Hayes DF, et al. Lack of correlation between gene variants in tamoxifen metabolizing enzymes with primary endpoints in the ATAC trial. Program and abstracts of the 33rd Annual San Antonio Breast Cancer Symposium, 9 December 2010.
- Leyland-Jones B, Regan MM, Bouzyk M, et al. Outcome according to CYP2D6 genotype among postmenopausal women with endocrine-responsive early invasive breast cancer randomized in the BIG 1-98 trial. Program and abstracts of the 33rd Annual San Antonio Breast Cancer Symposium, 9 December 2010.
- Roche PC, Suman VJ, Jenkins RB, et al. Concordance between local and central laboratory HER2 testing in the breast intergroup trial N9831. *J Natl Cancer Inst* 2002;94:855–857.
- Paik S, Bryant J, Tan-Chiu E, et al. Real-world performance of HER2 testing—National Surgical Adjuvant Breast and Bowel Project experience. *J Natl Cancer Inst* 2002;94:852–854.
- Wolff AC, Hammond ME, Schwartz JN, et al.; American Society of Clinical Oncology; College of American Pathologists. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *J Clin Oncol* 2007;25:118–145.
- Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group. Recommendations from the EGAPP Working Group: can tumor gene expression profiling improve outcomes in patients with breast cancer? *Genet Med* 2009;11(1):66–73.
- Zujewski JA, Kamin L. Trial assessing individualized options for treatment for breast cancer: the TAILORx trial. *Future Oncol* 2008;4:603–610.
- SWOG. RxPONDER trial will evaluate whether gene expression test can drive chemotherapy choice. <http://swog.org/visitors/newsletters/2011/04/index.asp?a=spotlight>. Accessed 16 September 2011.
- Wacholder S, Hartge P, Prentice R, et al. Performance of common genetic variants in breast-cancer risk models. *N Engl J Med* 2010;362:986–993.
- Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst* 1989;81:1879–1886.
- Gwinn M, Grossniklaus DA, Yu W, et al. Horizon scanning for new genomic tests. *Genet Med* 2011;13:161–165.
- Powell AE, Davies HT, Thomson RG. Using routine comparative data to assess the quality of health care: understanding and avoiding common pitfalls. *Qual Saf Health Care* 2003;12:122–128.
- Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol* 2005;58:323–337.
- Thorpe KE, Zwarenstein M, Oxman AD, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *J Clin Epidemiol* 2009;62:464–475.
- Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. *J Clin Epidemiol* 2009;62:499–505.
- Mohr PE, Tunis SR. Access with evidence development: the US experience. *Pharmacoeconomics* 2010;28:153–162.
- Trueman P, Grainger DL, Downs KE. Coverage with Evidence Development: applications and issues. *Int J Technol Assess Health Care* 2010;26:79–85.
- Centers for Medicare and Medicaid Services. Proposed Decision Memo for Pharmacogenomic Testing for Warfarin Response (CAG-00400N). <https://www.cms.gov/medicare-coverage-database/details/nca-proposed-decision-memo.aspx?NCAId=224&ver=15&NcaName=Pharmacogenomic+Testing+for+Warfarin+Response&NCId=333&ncdver=1&IsPopup=y&bc=AAAAAAAAIAAA&>. Accessed 16 September 2011.
- Eldridge S, Ashby D, Bennett C, Wakelin M, Feder G. Internal and external validity of cluster randomised trials: systematic review of recent trials. *BMJ* 2008;336:876–880.
- Varnell SP, Murray DM, Janega JB, Blitstein JL. Design and analysis of group-randomized trials: a review of recent practices. *Am J Public Health* 2004;94:393–399.

43. Mazor KM, Sabin JE, Boudreau D, et al. Cluster randomized trials: opportunities and barriers identified by leaders of eight health plans. *Med Care* 2007;45(10 suppl 2):S29–S37.
44. Qureshi N, Armstrong S, Saukko P, et al. Realising the potential of the family history in risk assessment and primary prevention of coronary heart disease in primary care: ADDFAM study protocol. *BMC Health Serv Res* 2009;9:184.
45. Emery J, Morris H, Goodchild R, et al. The GRAIDS Trial: a cluster randomised controlled trial of computer decision support for the management of familial cancer risk in primary care. *Br J Cancer* 2007;97:486–493.
46. O'Neill SM, Rubinstein WS, Wang C, et al.; Family Healthware Impact Trial group. Familial risk for common diseases in primary care: the Family Healthware Impact Trial. *Am J Prev Med* 2009;36:506–514.
47. Freidlin B, Jiang W, Simon R. The cross-validated adaptive signature design. *Clin Cancer Res* 2010;16:691–698.
48. Barker AD, Sigman CC, Kelloff GJ, Hylton NM, Berry DA, Esserman LJ. I-SPY 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. *Clin Pharmacol Ther* 2009;86:97–100.
49. Schneeweiss S. Developments in post-marketing comparative effectiveness research. *Clin Pharmacol Ther* 2007;82:143–156.
50. Habel LA, Shak S, Jacobs MK, et al. A population-based study of tumor gene expression and risk of breast cancer death among lymph node-negative patients. *Breast Cancer Res* 2006;8:R25.
51. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004;351:2817–2826.
52. Esteva FJ, Sahin AA, Cristofanilli M, et al. Prognostic role of a multigene reverse transcriptase-PCR assay in patients with node-negative breast cancer not receiving adjuvant systemic therapy. *Clin Cancer Res* 2005;11:3315–3319.
53. DeStefano F, Whitehead N, Lux LJ, Lohr KN. Infrastructure to monitor utilization and outcomes of gene-based applications: an assessment. (Prepared by RTI International DEcIDE Center under Contract No. HSA2902200500361.) AHRQ Publication No. 08-EHC012. Rockville, MD: Agency for Healthcare Research and Quality, May 2008.
54. Roth JA, Garrison LP Jr, Burke W, Ramsey SD, Carlson R, Veenstra DL. Stakeholder perspectives on a risk-benefit framework for genetic testing. *Public Health Genomics* 2011;14:59–67.
55. Basu A, Meltzer D. Modeling comparative effectiveness and the value of research. *Ann Intern Med* 2009;151:210–211.
56. Claxton KP, Sculpher MJ. Using value of information analysis to prioritise health research: some lessons from recent UK experience. *Pharmacoeconomics* 2006;24:1055–1068.
57. Gold MR, Siegel JE, Russell LB, Weinstein MC. *Cost-Effectiveness in Health and Medicine*. Oxford University Press: New York, 1996.
58. Grosse SD, Khoury MJ. What is the clinical utility of genetic testing? *Genet Med* 2006;8:448–450.
59. Payne K, Shabaruddin FH. Cost-effectiveness analysis in pharmacogenomics. *Pharmacogenomics* 2010;11:643–646.
60. Neumann PJ, Cohen JT, Hammitt JK, et al. Willingness-to-pay for predictive tests with no immediate treatment implications: a survey of US residents. *Health Econ* 2012;21:238–251.
61. Regier DA, Ryan M, Phimister E, Marra CA. Bayesian and classical estimation of mixed logit: an application to genetic testing. *J Health Econ* 2009;28:598–610.
62. Boote J, Barber R, Cooper C. Principles and indicators of successful consumer involvement in NHS research: results of a Delphi study and subgroup analysis. *Health Policy* 2006;75:280–297.
63. O'Haire C, McPheeters M., Nakamoto EK, et al. Methods for Engaging Stakeholders To Identify and Prioritize Future Research Needs. Methods Future Research Needs Report No. 4. (Prepared by the Oregon Evidence-based Practice Center and the Vanderbilt Evidence-based Practice Center under Contract No. 290-2007-10057-I.) AHRQ Publication No. 11-EHC044-EF. Agency for Healthcare Research and Quality: Rockville, MD. [http://www.effectivehealthcare.ahrq.gov/ehc/products/200/698/MFRNGuide04-Engaging\\_Stakeholders-6-10-2011.pdf](http://www.effectivehealthcare.ahrq.gov/ehc/products/200/698/MFRNGuide04-Engaging_Stakeholders-6-10-2011.pdf). Accessed 16 September 2011.
64. Hoffman A, Montgomery R, Aubry W, Tunis SR. How best to engage patients, doctors, and other stakeholders in designing comparative effectiveness studies. *Health Aff (Millwood)* 2010;29:1834–1841.
65. Helfand M, Tunis S, Whitlock EP, et al.; Methods Work Group of the National CTSA Strategic Goal Committee on Comparative Effectiveness Research. A CTSA agenda to advance methods for comparative effectiveness research. *Clin Transl Sci* 2011;4:188–198.
66. Foster MW, Mulvihill JJ, Sharp RR. Evaluating the utility of personal genomic information. *Genet Med* 2009;11:570–574.
67. Khoury MJ, Gwinn M, Yoon PW, Dowling N, Moore CA, Bradley L. The continuum of translation research in genomic medicine: how can we accelerate the appropriate integration of human genome discoveries into health care and disease prevention? *Genet Med* 2007;9:665–674.
68. Lord SJ, Irwig L, Simes RJ. When is measuring sensitivity and specificity sufficient to evaluate a diagnostic test, and when do we need randomized trials? *Ann Intern Med* 2006;144:850–855.
69. Rubin DB. On the limitations of comparative effectiveness research. *Stat Med* 2010;29:1991–1995; discussion 1996.
70. Selker HP, Strom BL, Ford DE, et al. White paper on CTSA consortium role in facilitating comparative effectiveness research: September 23, 2009 CTSA consortium strategic goal committee on comparative effectiveness research. *Clin Transl Sci* 2010;3:29–37.
71. Tunis SR, Benner J, McClellan M. Comparative effectiveness research: Policy context, methods development and research infrastructure. *Stat Med* 2010;29:1963–1976.
72. Tunis SR, Benner J, McClellan M. Response to comments on 'Comparative Effectiveness Research'. *Stat Med* 2010;29:1996–1997.