Title
Economic Perspectives on Personalized Health Care and Prevention

Journal
Forum for Health Economics & Policy, 16(2)

ISSN
2194-6191

Authors
Phillips, Kathryn A
Sakowski, Julie Ann
Liang, Su-Ying
et al.

Publication Date
2013-09-01

DOI
10.1515/fhep-2013-0010

Peer reviewed

DE GRUYTER DOI 10.1515/fhep-2013-0010 — Forum for Health Economics and Policy 2013; 16(2): S23–S52

Kathryn A. Phillips*, Julie Ann Sakowski, Su-Ying Liang and Ninez A. Ponce

Economic Perspectives on Personalized Health Care and Prevention

Abstract: The objective of this paper is to provide an overview of economic evaluation of personalized medicine, focusing particularly on the use of cost-effectiveness analysis and other methods of valuation. We draw on insights from the literature and our work at the University of California, San Francisco Center for Translational and Policy Research on Personalized Medicine (TRANSPERS). We begin with a discussion of why personalized medicine is of interest and challenges to adoption, whether personalized medicine is different enough to require different evaluation approaches, and what is known about the economics of personalized medicine. We then discuss insights from TRANSPERS research and six areas for future research:

1. Develop and Apply Multiple Methods of Assessing Value
2. Identify Key Factors in Determining the Value of Personalized Medicine
3. Use Real World Perspectives in Economic Analyses
4. Consider Patient Heterogeneity and Diverse Populations in Economic Analyses
5. Prepare for Upcoming Challenges of Assessing Value of Emerging Technologies
6. Incorporate Behavioral Economics into Value Assessments

Keywords: costs and cost-effectiveness analysis; economics; health policy; individualized medicine; personalized medicine; preferences measurement.

***Corresponding author: Kathryn A. Phillips,** University of California, San Francisco – Department of Clinical Pharmacy, 3333 California Street Box 0613, San Francisco, CA 94143, USA, Phone: +(415) 502-8271, e-mail: phillipsk@pharmacy.ucsf.edu
Julie Ann Sakowski: Associate Professor of Health Economics, Department of Clinical Pharmacy, UCSF and Executive Director, UCSF Center for Translational and Policy Research on Personalized Medicine
Su-Ying Liang: Research Professor of Health Economics, UCSF Center for Translational and Policy Research on Personalized Medicine and Department of Clinical Pharmacy UCSF; and Research Economist, Department of Health Policy Research, Palo Alto Medical Foundation Research Institute
Ninez A. Ponce: Associate Professor, UCLA Fielding School of Public Health Department of Health Policy and Management

1 Introduction

The objective of this paper is to provide an overview of economic evaluation of personalized medicine (PM), focusing particularly on the use of cost-effectiveness analysis (CEA) and other methods of valuation.

We begin with definitions to clarify the landscape and scope of this paper. The term "personalized medicine" generally is used to refer to targeting of health care based on genetic information, although it also can be thought of more broadly as targeting based on any characteristic such as patient preferences. Recently, a National Academy of Sciences report used the term "precision medicine" – the use of genomic, epigenomic, exposure, and other data to define individual patterns of disease, potentially leading to better individual treatment. It was felt by the committee that this definition conveys a more accurate image of diagnosis that is person-centered and multifaceted (National Research Council 2011). PubMed has a category for "individualized medicine," defined as "therapeutic approaches tailoring therapy for genetically defined subgroups of patients." In earlier years, the more commonly used term was "pharmacogenomics," which is a narrower concept, defined as the analysis of the effect of genomics – in particular, genetic variation (polymorphisms) – on drug response.

Economic evaluation can encompass a wide range of topics and methods, including some that would be more narrowly considered as economic analyses based on standard welfare economic theory as well as others derived from fields such as decision analysis, operations research, and behavioral economics. The most commonly used form of economic evaluation is CEA (Gold et al. 1996). Related approaches include cost-benefit analysis, decision analysis, stated choice methods/conjoint analysis/discrete choice experiments and willingness-to-pay, budget impact analysis, burden of illness studies, value of information analysis, and multi-criteria decision analysis.

Regardless of how one defines PM, it is clear that the trend towards greater targeting of interventions to sub-populations is inevitable given our growing understanding of disease. For example, there have been over 400 new genetic tests since 2010 (Khoury et al. 2012a), and an Agency for Healthcare Research and Quality (AHRQ) horizon scan of genetic tests for cancer found that over 100 tests are clinically available – a 67% increase in 4 years (Raman et al. 2011). It is hoped that PM will result in higher-quality, lower-cost health care because of opportunities to offer patients therapies that are more effective for them and avoid treatments that will not be safe or effective.

Several reviews have provided overviews of economic opportunities and challenges for economics of PM using a variety of perspectives. Deverka et al. (2010)

noted that there is great potential for genetic testing to be cost-effective and even cost-saving, although they also noted that its ability to do so will depend on many factors. A study by McKinsey Consulting examined the major barriers to adoption of PM from a business perspective (Davis et al. 2009). They concluded that there is poor alignment of incentives across key business stakeholders (e.g., payers, providers, pharmaceutical and biotech companies, diagnostic companies). For example, they noted that a test that reduces cancer treatments may reduce physician revenues, a test that increases the number of patients using treatments with marginal health benefits may increase health plan costs, and some applications do not offer potential for economic gains from development of companion tests.

In a commentary, Katrina Armstrong noted that genomics could potentially "bend the cost curve" but that several steps would need to be taken to enable it to do so, including effective clinical decision support, information systems, guidelines about how to manage genomic information unrelated to the clinical question of interest, and testing costs that are lower than the cost of the interventions (Armstrong 2012). However, despite the conceptual appeal of this assertion, it is also possible that PM will simply increase costs without increasing benefits. Difficult decisions will have to be made about how to assess the value of these technologies, which technologies will be adopted, and who will pay for them – all areas that economics can address.

PM is also of interest because of the increased emphasis on patient-centered care that takes into account patient variability, including genetic differences. The authorizing legislation for the Patient-Centered Outcomes Research Institute (PCORI) states that research shall be designed to take into account potential for differences in the effectiveness of health care as used with various *subpopulations*. Genetic factors may cause treatment effects to vary across individuals, along with clinical factors, personal and behavioral factors, and the health care delivery context.

Despite the conceptual appeal of personalized medicine, there are many barriers to its adoption. Key challenges include (1) negotiating shifting industry paradigms; (2) balancing innovation and regulation; (3) building the evidence base; and (4) determining value and reimbursement. Figure 1 provides a conceptual framework by placing economic analyses within a larger context. This framework notes that the use of genomic information is part of a translational continuum – from basic research, to clinical research, to policy research – that determines adoption and health and economic outcomes. Key determinants of adoption and outcomes include utilization, tradeoffs as defined by preferences and by economic benefits/costs, and evidence. Economics permeates each of these areas, with a central focus on preferences and demand based on the utility obtained from consuming goods and services.

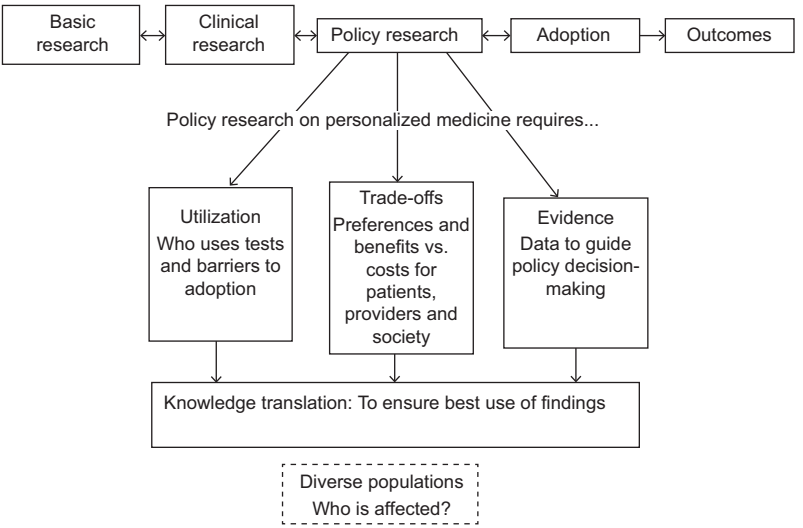

Figure 1 Conceptual Framework for Personalized Medicine Translation.

2 Does Personalized Medicine Require Different Evaluation Approaches?

It is clear that, historically, the evaluation of diagnostics has been different from that of drugs (Phillips et al. 2006) – and most PM interventions include diagnostics. The regulatory pathways, the measures of clinical utility, and the number and scope of economic analyses have been different for diagnostics. However, one reason is that diagnostics have typically been low-cost, low-risk interventions and thus the level of scrutiny has been less. With the advent of diagnostics based on genetic information, that situation has changed such that PM diagnostics are often high-cost, higher-risk interventions that guide clinical care, and thus evaluation approaches may be shifting.

In general, the literature suggests that standard methods can be applied to PM; however, there are a number of complexities of PM that may require different approaches to economic evaluation (Phillips et al. 2006). The field of genetics has historically not been a focus of economists or health services researchers, and understanding these technologies can require more interdisciplinary work across basic, clinical, and social scientists. Laboratories and coding systems play a large role in PM adoption, adding another layer of complexity. The Food and Drug Administration (FDA) does not directly regulate most diagnostics and thus does not assess their clinical utility, often resulting in less information available

on their efficacy and effectiveness. Emerging PM tests are often complex technologies – based on algorithms and panels of genes – and they may be used for multiple purposes, so they can be harder to assess than single gene tests. Genomic technologies based on inherited mutations such as BRCA testing may impact family members, requiring more complex models that incorporate both the initial patients and their families.

Two key characteristics of PM that require more complex analyses and are often poorly understood are (1) diagnostics can be characterized by varying degrees of analytic validity, clinical validity, and clinical utility, and (2) diagnostics provide information rather than directly changing outcomes. A useful framework for understanding the evaluation of genetic tests and the role of analytic validity, clinical validity, and clinical utility is the framework developed by the Centers for Disease Control and Prevention (http://www.cdc.gov/genomics/gtesting/ACCE/index.htm). The ACCE framework uses key criteria for evaluating a genetic test – analytic validity, clinical validity, clinical utility and associated ethical, legal and social implications:

- Analytic validity – How accurately and reliably the test measures the genotype of interest.
- Clinical validity – How consistently and accurately the test detects or predicts the intermediate or final outcomes of interest.
- Clinical utility – How likely the test is to significantly improve patient outcomes.

Economic evaluations often confuse analytic and clinical validity; for example, they will model a test as 100% accurate because its analytic validity is 100% – but its clinical validity is typically less than that. Thus, determining the accuracy of tests can be complicated. In addition to the conceptual challenges, accuracy can be difficult to assess because it will also vary depending on the laboratory where it is conducted.

Another key characteristic of PM that increases the complexity of analyses is that diagnostic tests produce information rather than direct value. Thus, economic evaluations need to consider how that information impacts treatment decisions and outcomes, and both the test and intervention need to be considered simultaneously. There may be a variety of testing options, different pathways linking tests to therapies, and tests that apply to several drugs. Figure 2 shows that the actual effectiveness and cost-effectiveness of diagnostics depend upon whether a test is ordered, whether the test accurately identifies who should get treatment, and whether patients and providers act in accordance with test results. However, economic analyses often make simplifying assumptions about these factors that do not reflect real-world test characteristics and behaviors.

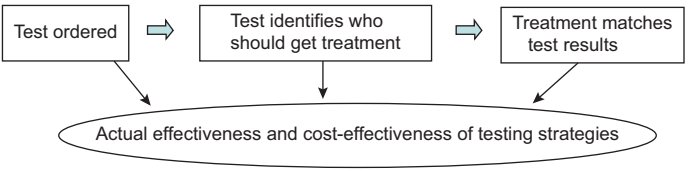

Figure 2 The Sequence of Events for Diagnostic Testing.
Source: Phillips et al. 2009.

3 What is Known about the Economic Evaluation of Personalized Medicine?

Early studies focused on reviews of the societal and economic implications of personalized medicine, while more recently a number of empirical analyses and cost-effectiveness analyses have been published. Examples include an early review of societal and economic implications (Phillips et al. 2000); a primer on applying CEA and cost-benefit analysis to pharmacogenomics (Phillips et al. 2003); and a proposal for a resource allocation framework from a population perspective using CEA and cost-of-illness studies (Phillips and Van Bebber 2005). Many others have since published conceptual reviews of the field. For example, Khoury and colleagues laid out a population approach to precision medicine that includes economic analyses and expertise (Khoury et al. 2012b).

One of the earliest empirical studies concerned the potential role of pharmacogenomics in reducing adverse drug reactions (Phillips et al. 2001). It suggested that drug therapy based on individuals' genetic makeups may result in a clinically important reduction in adverse outcomes. This study required the integration of two research areas: basic and clinical research on genetics and population science on adverse drug reactions and quality of care – and a team comprised of basic and clinical scientists and social scientists.

One of the earliest reviews of pharmacogenomics identified only 11 economic evaluation studies, predominantly examining inherited mutations (Phillips and Van Bebber 2004). Since then, there have been several structured reviews to synthesize and characterize the body of evidence on the economic value of clinical genomic applications (Giacomini et al. 2003; Carlson et al. 2005; Vegter et al. 2008; Beaulieu et al. 2010; Wong et al. 2010; Djalalov et al. 2011). These reviews have aggregated and drawn conclusions about the cost-effectiveness of various types of PM applications, assessed the quality of the available evidence, and proposed methodological standards that should be put in place to ensure that economic evaluations are producing the information

needed for decision-making. Our assessment of the conclusions from these reviews is:

- Economic evaluations have only been conducted on a fraction of the available genomic tests and on limited topics such as hepatitis C and breast cancer. However the evidence base is growing and at an accelerating rate.
- Studies vary in terms of how they define the topic of focus and how they use different search terms, thus making it difficult to assess the evidence base as a whole.
- A majority of the economic evaluations identified in these reviews are CEAs followed by cost-utility analyses, with fewer cost-benefit analyses and cost minimization analyses.
- A majority of studies included in these reviews found PM interventions to have favorable CEA ratios, although fewer interventions were cost-saving. Beaulieu et al. (2010) noted that parameters that significantly impacted cost-effectiveness and that should be included are biomarker prevalence, population ethnicity, pharmacogenomic treatment effect, and cost of genomic data collection and analysis.
- Assessments of the quality of the CEAs suggest general adherence to quality guidelines with some exceptions, and that quality is improving over time. Wong et al. (2010) concluded that the quality of studies could be improved by careful evaluation of the clinical validity and potential clinical utility of proposed biomarkers.

There are two major challenges to assessing the economic valuations of PM. First, there are no straightforward means by which to search for these studies in PubMed. There is no Major Exact Subject Heading (MeSH) in PubMed for "personalized medicine." Relevant studies have been coded with a variety of terms such as "pharmacogenetics" and "genetic testing." In 2010, a MeSH term for "individualized medicine" was added to PubMed, including a subheading for "economics." But only a fraction of the published studies have been coded so far using this term. Second, there is currently no consolidated source of information on the economic value of PM that includes a range of evaluation methods. The evidence base in the published literature on PM remains scattered thus making it difficult to draw conclusions about the value of PM in general and identify its most appropriate uses and applications. There have been recent efforts to aggregate PM evidence, such as the National Institutes of Health (NIH) Genetic Test Registry, the Genomic Applications in Practice and Prevention Network (GAPPNet), and the Pharmacogenomics Knowledgebase (PharmGKB). A limitation of these consolidated resources, however, is that they focus on the clinical and analytic characteristics of the genetic test rather than economic information.

Since 2008, the University of California, San Francisco Center for Translational and Policy Research on Personalized Medicine (TRANSPERS) has conducted interconnected projects on utilization, preferences, cost-effectiveness, regulation, reimbursement, evidence development, diverse populations, and health policies. Findings emerging from these studies are summarized in this section.

3.1 Lack of Evidence on Personalized Medicine in Real-World Settings

Two TRANSPERS studies described gaps in the PM evidence base and how these could be addressed (Phillips 2008; Phillips et al. 2009). Human Epidermal Growth Factor Receptor 2 (HER2) testing to target trastuzumab treatment for patients with breast cancer is perhaps the best known example of testing to target treatment. HER2 testing determines which patients overexpress the gene HER2; for those 20–30% of patients, trastuzumab is highly effective. Trastuzumab and an accompanying test were approved in 1998 for patients with metastatic breast cancer and its use was expanded to patients with early-stage breast cancer after 2005. HER2 testing is now recommended for all patients with invasive breast cancer, and only patients with positive test results are recommended for trastuzumab treatment. There is no consensus about optimal testing methods. Guidelines recommend using either immunohistochemistry (IHC), with indeterminate results confirmed by fluorescence in situ hybridization (FISH), or FISH alone, to determine HER2 status. Although FISH is a better predictor of response to treatment, immunohistochemistry costs substantially less and is more easily performed in community laboratories.

There was little available information on the actual use of HER2 testing in clinical practice, including whether all eligible patients receive testing, the impact on follow-up testing and changes in treatment decisions, and access to testing by underserved populations. At the time of our analyses, there was little or no information on whether women who were uninsured, on Medicaid, or minorities received testing and the extent to which women received IHC or FISH or both. There was evidence to suggest that a large percentage of tests were inaccurate and that women who tested negative were still getting treatment – and conversely, women who tested positive did not always get treatment – especially lower income women. We found that most of the CEAs conducted had assumed perfect testing and use of test results – and thus may have overstated the actual cost-effectiveness of testing and treatment in clinical practice.

Several recommendations to improve the evidence base emerged from this study and others. A first step is to document and disseminate gaps in knowledge about actual clinical practices. In the case of HER2 testing, clinicians and insurers often were surprised to hear about the evidence gaps.

Second, there is a need to standardize billing codes, laboratory procedures and documentation, claims information, and medical records for PM testing to enable better analyses of access and utilization. It is often impossible to identify the use of testing in administrative databases because of coding issues. For example, without chart review, the use of IHC and FISH for HER2 detection cannot be distinguished reliably from the same types of tests performed for other indications. Administrative claims and medical records may not match; we found a mismatch for HER2 testing claims and records about 25% of the time (Liang et al. 2011).

A third need is to examine the impact of policies intended to reduce data gaps and encourage more appropriate use of interventions. One example is the policy change implemented by United Healthcare in 2006, after it was found that many patients prescribed trastuzumab had missing or negative test results (L. Newcomer, personal communication, August 14, 2008). Its policy required clinicians to submit documentation of a positive HER2 test result with the first trastuzumab claim. The rate of submitted claims for trastuzumab decreased after the policy implementation, suggesting that it may reduce inappropriate use. Because of this policy and other efforts to improve laboratory procedures, the plan has more information about why errors are made in HER2 testing and has been able to implement quality improvement programs. Although there are potential disadvantages to such policies, they may warrant consideration for some technologies.

Lastly, we need to develop creative approaches to obtaining evidence. Currently, few databases link testing, test results, treatment, and outcomes, and no system in the United States regularly monitors tests after their adoption. Work being done by PCORI to enhance the use of observational databases, including more sophistical statistical approaches and greater linkages, should improve our ability to assess real-world practice. Linking groups such as payers with researchers is also a fruitful approach; TRANSPERS has worked with several health plans and their foundations and, most recently, the Health Care Cost Institute has aggregated claims data from several of the nation's largest insurers.

3.2 Will Personalized Medicine be Cost-saving or More Cost-Effective than Alternatives?

Our work to date on Lynch syndrome and gene expression profiling for breast cancer suggests that PM interventions may be cost-effective – given certain

assumptions – but that one cannot assume that PM will be cost-saving or reduce costs of care.

Lynch syndrome is the most common genetic cause of colorectal cancer. Testing for the syndrome can identify family members who will benefit from increased interventions for cancer prevention and early treatment. We conducted a CEA of testing for Lynch syndrome, comparing various testing strategies (Ladabaum et al. 2011). Among tumor-testing strategies, IHC followed by BRAF gene mutation testing was preferred, with an incremental cost-effectiveness ratio of $36,200 per life-year gained. We concluded that testing for Lynch syndrome could yield substantial benefits at acceptable costs, particularly for women who begin regular screening and have risk-reducing surgery.

Gene expression profiling (GEP) testing for breast cancer provides information about cancer recurrence and chemotherapy decisions. Several companies offer GEP testing for breast cancer, but the Onco*type*DX™ breast cancer assay is the most commonly used test, and is covered by insurance in the United States despite a cost of almost $4000 per test. We conducted a CEA of GEP for breast cancer, comparing the most commonly used test (Onco*type*DX™) with a common clinical algorithm developed by the National Comprehensive Cancer Network (NCCN). Preliminary results suggest that using GEP may be more cost-effective than relying on clinical algorithms but that there is a great deal of uncertainty around these results, and results are heavily impacted by assumed use of chemotherapy following testing (Marshall et al. 2012).

In addition, we conducted an analysis to examine the association between GEP testing and use of chemotherapy, serious chemotherapy-related adverse effects, and total charges during the 12 months following diagnosis (Haas et al. 2011). Data were from administrative claims and medical records from a national health plan. Our findings suggest that GEP testing was most commonly used in women at moderate (versus low or high) clinical risk of recurrence (52% vs. 25% of low-risk women and 5.5% of high-risk). While GEP testing was associated with an overall decrease in adjuvant chemotherapy, we did not find differences in serious chemotherapy-associated adverse events or charges during the 12 months following diagnosis.

3.3 The Method of Targeting

Assessing the impact of a targeted intervention on costs and health outcomes requires explicit consideration of the method of targeting – but we found in a review of breast cancer studies that few studies explicitly evaluated the relationships among the method of targeting, the accuracy of the targeting test, and outcomes of the targeted intervention (Elkin et al. 2011). Studies that did evaluate the

method of targeting found that characteristics of targeting tests had a substantial impact on outcomes.

Nearly all diagnostic tests give an inherently continuous result that is categorized as a basis for action, and nearly all tests must be considered imperfect predictors of a true state, rather than certain indicators of the truth. When the outcome of a targeted intervention depends on the presence or absence of the target, test results are generally dichotomized, based on some threshold applied to the underlying continuous result. In these cases, economic evaluation of a targeted strategy requires explicit consideration of test performance relative to a gold standard and how the performance characteristics of the test in the population of interest – sensitivity, specificity, positive predictive value, and negative predictive value – influence the use and outcomes of the intervention.

A CEA restricted to a cohort with positive test results captures only part of the full range of health and economic impacts of implementing a targeted intervention because it ignores negative test results and their consequences (Figure 3). When the targeting test is very expensive, a CEA that ignores individuals with negative test results would exclude these costs. Such an analysis would also exclude the negative health consequences of failing to give the intervention to those with a false-negative result – individuals who truly have the target.

Second, economic evaluation of risk-targeted interventions requires explicit consideration of the threshold risk criterion. Figure 4 shows hypothetical distributions of the predicted risk of disease recurrence in two groups of women treated for early-stage breast cancer: those who will, in fact, experience a recurrence and those who will not. In Panel A, the threshold is relatively strict; the

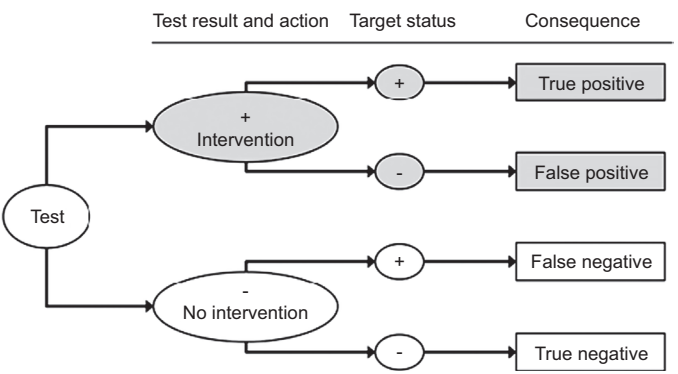

Figure 3 Possible Test Results in an Economic Evaluation of a Targeted Intervention.
Source: Elkin et al. 2011.

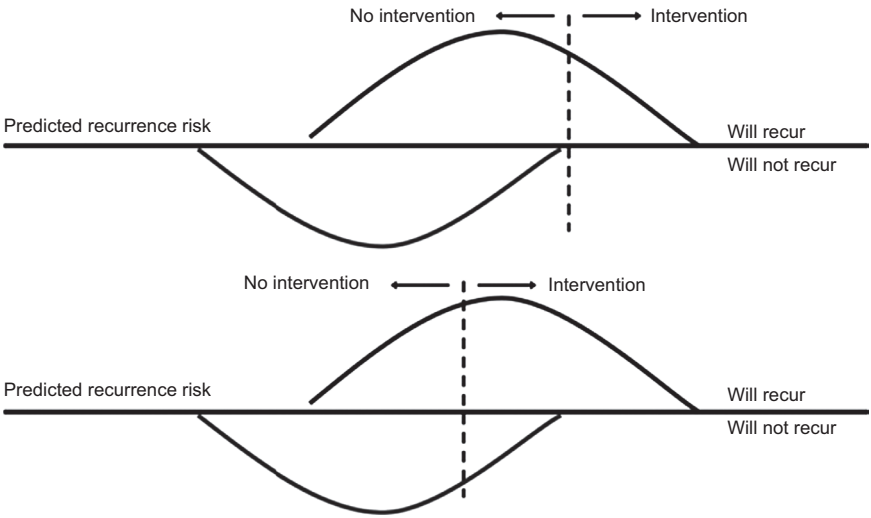

Figure 4 Thresholds for a Targeted Intervention by Predicted Risk and Actual Outcome. Source: Elkin et al. 2011.

intervention is given to a small proportion of patients, all of whom would have had a disease recurrence in the absence of the intervention. When the criterion is more lenient and individuals with a lower predicted risk are eligible for the intervention (Panel B), more patients who would have recurred receive the intervention but so too will some women who never would have had a disease recurrence. The tradeoffs associated with the threshold risk criterion – and the resulting costs, risks, and benefits of the test and intervention – will influence the cost-effectiveness of a risk-targeted strategy.

3.4 Behavior of Family Members

Behavior of family members can be a key factor in determining the value of PM for inherited mutations. In the CEA of Lynch syndrome screening, we found that it is especially important to identify and counsel relatives of the persons being tested because the benefits of screening cannot be realized otherwise (Kreft 1995). We found that the number of relatives tested per proband (person presenting with Lynch) was a critical determinant of both effectiveness and cost-effectiveness, with testing of 3 to 4 relatives required for most strategies to meet a threshold of $50,000 per life-year gained. However, there is a lack of empirical evidence on whether family members will change their behaviors.

3.5 Patient and Provider Preferences

Patient and provider preferences may affect the cost-effectiveness of personalized medicine. Our conjoint analysis found that individuals have a relatively high value for Lynch syndrome screening – if such screening is private and accurate (Knight 2011). Each choice task included two hypothetical genetic test alternatives (with varying attributes for accuracy, privacy, and cost), a no-test option, and a hypothetical level of colorectal cancer risk. We found that privacy was the most valued attribute. Most participants would have genetic testing to reduce the risk of dying from colorectal cancer in the best scenario (no false-negatives, results disclosed to primary care physician), while only 41% would have genetic testing in the worst case (20% false-negatives, results disclosed to insurance company). The overall monetary value of testing relative to no testing was $622.

Our utilities study and cost-utility study also suggested that patient preferences may play a large role in uptake of testing, and thus impact the cost-effectiveness of genetic testing. In our time tradeoff study, we found that preferences for the potential outcomes of testing vary substantially, calling into question the extent to which patients would avail themselves of such testing if it were offered (Kuppermann et al. 2012). Less than half of the sample assigned higher scores to undergoing Lynch testing and receiving negative results versus foregoing testing. This suggests that knowing that one does not carry a Lynch syndrome-causing mutation may *not* be viewed as a net gain to many of the people to whom current guidelines are directed.

In our cost-utility analysis of Lynch syndrome screening, we found that cost-effectiveness of genetic testing decreased when quality-of-life considerations were included, although testing was still relatively cost-effective (Wang et al. 2012). The duration and magnitude of decreases in quality-of-life after decisions related to testing and surgeries were key determinants of the cost-effectiveness of screening. Similarly, in an ongoing study of the cost-effectiveness of gene expression profiling for breast cancer, we found that patient and physician preferences on treatment choices – and thus the proportion of women receiving chemotherapy – have a large impact on cost-effectiveness of testing (Marshall et al. 2012).

Provider preferences are also likely to be important to cost-effectiveness. We found in a study of colorectal cancer screening that physicians were more likely to underestimate how much patients valued screening, which could affect screening rates because physicians may be less likely to recommend screening if they believe that their patients do not value it (Marshall et al. 2009).

3.6 Assessing the Value of Personalized Medicine

Payers struggle with how to assess the value of PM for making coverage and reimbursement decisions. We have conducted several studies exploring how payers make coverage and reimbursement decisions on PM technologies (Trosman et al. 2010, 2011). Many of these studies have emerged from interviews with our Evidence and Reimbursement Policy Council, which include senior executives from the seven largest United States health insurance plans, leading regional plans, and pharmacy benefits management companies as well as thought leaders with industry, government, and Medicare perspectives.

Of particular relevance is a study of the health technology assessment guidelines used by payers in decision making (Trosman et al. 2011). We found that payers typically used one or more of seven health technology assessment guidelines:

- Blue Cross Blue Shield Technology Evaluation Center (BCBS TEC)
- Emergency Care Research Institute (ECRI)
- Evaluation of Genomic Applications in Practice and Prevention (EGAPP)
- Institute for Clinical and Economic Review (ICER)
- United States Preventive Services Task Force (USPSTF)
- UpToDate
- Hayes

Payers varied greatly in how many guidelines they used, with some payers using all or most of them while others used one or two. Guidelines varied as to their inclusion of costs and economic analyses, from minimal to a central focus. Payers noted that they expect to use cost-effectiveness criteria in the future and thus the groundwork should be laid now. Payers generally stated that they do not explicitly use economic analyses in their decision making – although they expect to do so in the future. As one payer stated, "We don't use cost-effectiveness in decisions today, but in the future state, it's critical that we do that and get some agreement on how we measure the value and cost-effectiveness. The groundwork needs to be lain but it's not there today" (Trosman et al. 2011: p. 21s).

Payers noted that they focused first on clinical utility in making decisions – but when this was uncertain, they wanted information on other nonclinical factors such as cost-effectiveness. However, payers felt that the guidelines were lacking in information on cost-effectiveness – 82% of payers stated that this was a shortcoming in health technology assessment reviews.

3.7 Diffusion of Personalized Medicine

Diffusion of PM appears to lag for certain subgroups, with implications for utilization and cost-effectiveness. For diverse populations, our findings document racial/ethnic and social gaps in access to and use of PM diagnostics. One reason may be because a perception of one's family history risk may vary by race/ethnicity (Ponce et al. 2012). On the supply side, PM diagnostics, such as GEP to inform use of adjuvant chemotherapy for early-stage breast cancer, have not been adequately validated in minority populations (Odierna et al. 2011). Public payers tend to be less generous in BRCA gene mutation testing than private payers, despite professional guidelines establishing BRCA testing as standard of care (Wang et al. 2011). Yet even among women with the same private insurance coverage, use of GEP was lowest for low-income women (Haas et al. 2011). Taken together, our studies suggest that socioeconomically vulnerable subgroups may lag behind in harnessing the benefits of PM, and the barriers stem from both preferences and supply-side policies.

4 Opportunities, Challenges, and Research Questions

4.1 Develop and Apply Multiple Methods of Assessing Value

The most commonly-used approach to measuring the economic value of health care interventions, including PM, is CEA. Furthermore, a particular type of CEA – cost-utility analysis (CUA) using quality-adjusted life years, is the recommended approach for what has been termed the "reference case" for comparability across studies (Gold et al. 1996). Despite the usefulness of CEA and CUA, there are a number of political and methodological barriers to using these methods, and thus the field can move forward by using a broader conception of "value" and wider use of alternative methods for assessing value.

Many commentators have noted that decision-makers are often reluctant to use CEA because of concerns about the appearance of rationing care based on costs and concerns about how such analyses are conducted (Neumann 2005). These concerns are highlighted by the restriction on use of cost-effectiveness analyses in the legislation establishing PCORI. PCORI is prohibited from using "dollars-per-quality adjusted life year (or similar measure that discounts the value of a life because of an individual's disability) as a threshold to establish

what type of health care is cost effective or recommended." As noted by PCORI Executive Director Joe Selby, "You can take it to the bank that PCORI will never do a cost-effectiveness analysis" (See http://insidehealthpolicy.com/Inside-Health-General/Public-Content/pcori-head-vows-not-to-do-cost-effectiveness-studies-but-notes-gray-areas/menu-id-869.html). However, he did note that costs are intertwined with health outcomes and thus may be considered as relevant. Stakeholder groups have noted discomfort with using quality-adjusted life years (QALYs) because they do not capture all relevant factors and because of concerns about the methods used to obtain them (Roth et al. 2011). Furthermore, Greenberg and Neumann (2011) found that adjusting life years for quality of life does not substantively affect cost per life year ratios, at least for cancer-related interventions.

These concerns illustrate the need for a broader conception of value and a range of metrics for its assessment. In lay terms, value is simply the importance of something. The economic concept of value further takes into account that value is relative – that it is based on a change in utility relative to a change in consumption – and that it involves tradeoffs with other goods or services. With this broader conception of value as a starting point, we can consider various approaches to measuring value, with the appropriate approach dependent on the question to be addressed and the relevant audience.

In addition to the various standard methods of valuation (cost-minimization analysis, cost-consequence analysis, CEA, and cost-benefit analysis), there are also various other approaches to measuring value that have emerged in recent years and that may be useful for evaluating PM. We discuss in this section several types that may prove relevant to economic evaluation of PM.

4.2 Resource Allocation Frameworks

Valuation methods such as CEA compare one intervention to another to examine incremental costs and effects, that is, they look at what is the "bang for the buck." However, CEA does not address the broader questions of where PM could have the greatest impact (Phillips et al. 2005). A PM test could be relatively cost-effective but have little impact on population health or be unaffordable within a budget-constrained environment.

Two resource allocation frameworks can be useful in evaluating these questions:

- Burden of illness (also called cost of illness) studies examine the total costs incurred by society due to a specific disease.

– Budget impact studies (BIA) estimate the financial consequences of adoption and diffusion of a new health care intervention within a specific health care setting or system given resource constraints. It has been recommended that a BIA accompany a CEA when payers are considering a new intervention since it is possible that a PM technology is efficient based on the CEA but not affordable based on the BIA due to the expected adoption and diffusion of the technology. (Mauskopf et al. 2007).

Budget impact and burden of illness studies shift the focus to the actual impact of interventions in the real world. Furthermore, their focus on affordability and population impact may be more politically palatable than a focus on cost-effectiveness.

To illustrate, we use the example of CYP2D6, which is one of the most studied drug-metabolizing enzymes that has been estimated to be responsible for metabolizing 25% of drugs, including commonly-used ones such as anti-depressants, beta-blockers, and codeine (See http://youscript.com/healthcare-professionals/what-is-youscript/pharmacogenetic-testing/cytochrome-p450-2d6-genotyping). Because CYP2D6 metabolizes an array of commonly used drugs, testing for CYP2D6 variants associated with slow or fast drug metabolism is likely to have implications not only for current drug utilization but also for future drug utilization, because test results can be used over a lifetime. Testing will also have implications for total budgets because of the potential for widespread testing and follow-up.

Key questions in a resource allocation framework are: What are the sizes of the relevant populations? What are the costs associated with those populations? What is known about the association of genetic variation with drug metabolism, response and clinical outcomes? Table 1 illustrates measures for addressing these questions using the example of CYP2D6.

4.3 Value of Information Analysis

Value of information analysis (VOI) offers a formal approach to deciding when and what types of data to collect. Formal use of decision analysis and VOI analysis can help determine whether an intervention should be adopted, whether additional evidence to further inform that decision is worth gathering, and what kind of information is of greatest value.

The use of VOI has recently been highlighted because of its emphasis in PCORI priorities. The Methodology Committee's draft report notes that VOI is a valuable tool for determining PCORI's funding priorities and they name it one

Table 1 Summary of Measures in a Pharmacogenetics Resource-Allocation Framework.

Relevant measure	Description	CYP2D6 example
Relevant populations		
Mutation prevalence	Measure of the size of the population in which testing could have an impact on outcomes	Prevalence of individuals with slow or rapid metabolism due to *CYP2D6* variant alleles
Drug utilization	Measure of the size of the population that could be tested	Utilization of drugs metabolized by CYP2D6
Prevalence of condition for which drug is used	Another measure of the size of the population that could be tested, but which includes individuals who are untreated or treated with another drug but for whom testing might be relevant	Prevalence for primary indications of drugs metabolized by CYP2D6
Relevant costs		
Drug expenditures	Measure of the potential outcomes of testing because testing could change the utilization of drugs	Expenditures on drugs metabolized by CYP2D6
Condition expenditures	Measure of the potential outcomes of testing because testing could change disease costs	Prevalence for primary indications of drugs metabolized by CYP2D6
Association of genetic variation		
Mutation effect on drug outcomes	Measure of the potential impact of testing because mutations must be associated with drug metabolism, drug response and clinical outcomes in order for testing to have an impact	Relationship of *CYP2D6* variant alleles to variation in metabolism, drug response and clinical outcomes

Source: Phillips et al. (2005). Measuring the value of pharmacogenomics. Nat Rev Drug Discov 4: 500–510.

of the four key components to a framework for establishing research priorities (Myers et al. 2012). However, lack of familiarity has kept this method from being used as widely as it could be.

VOI also provides a useful framework for setting priorities for further research into CEA. An example of this approach applied to PM is a cost-utility study of gene

expression profiling for breast cancer (Hall et al. 2012). The VOI analysis was able to show that the most important recommendation for further cost-effectiveness research is for future researchers to collect additional retrospective or prospective information on recurrence rates.

4.4 Stated Choice Methods and Willingness-to-pay

Stated choice methods (often called conjoint analysis or discrete choice experiments) measure preferences for constructed alternatives (Phillips et al. 2002). They are designed to provide information about individuals' willingness to accept tradeoffs among features of multi-attribute products. The preferred alternative is chosen by respondents among alternative scenarios described by multiple factors. Based on the choices made by the respondent for a series of pairs of hypothetical alternatives that systematically vary combinations of beneficial and harmful outcomes, we can estimate marginal rates of substitution (tradeoffs) among the factors. If cost is included as one of the attributes, willingness-to-pay can be estimated.

Willingness-to-pay could be a particularly important measure for evaluating PM because it may better reflect the value of information per se and the process of care, compared to other measures (Bridges et al. 2011). Studies have shown that information may have value for individuals even if it is not actionable, such as information on Alzheimer's risk (Neumann et al. 2012). However, there is wide debate over whether it is appropriate for payers to consider "intrinsic" value in coverage and reimbursement decisions because consumer demand for information is much higher than our ability to pay for tests to obtain information that is not actionable. There are also concerns that willingness-to-pay studies may provide an inflated value because they do not adequately reflect budget constraints – that even though survey respondents are told to consider the true cost to them (without insurance and relative to other goods and services) they are not able to do so given that most people have little sense of the true cost of health care they consume. There are also dilemmas about how to incorporate willingness-to-pay into value analyses. It can be used in cost-benefit analyses, but few of these are done. There have been attempts to better define how willingness-to-pay can be used in CEAs, but this remains a challenge (Polsky 2005).

4.5 Multi-Criteria Decision Analysis Including Risk-benefit Frameworks

There are a number of different systematic approaches for quantifying and comparing benefits and risks, often called "multi-criteria decision analysis" (Guo et al.

2010). The objective of these approaches is to promote evidence-based decision-making by organizing and synthesizing information and incorporating stakeholder perspectives. Such methods can help decision makers make more informed choices when faced with complex decisions involving several dimensions.

Two risk-benefit frameworks relevant to PM that might be considered as falling into this broad category are: (1) A risk-benefit framework for genetic testing that compares the certainty of evidence of value with the risk-benefit ratio – favorable, neutral, unfavorable (Veenstra et al. 2010), and (2) the Institute for Clinical and Economic Review (ICER) Integrated Evidence Rating™, which speaks both to comparative clinical effectiveness and comparative value and provides payers with a tangible tool uniquely designed to support value-based benefit designs, reimbursement strategies, and coverage policies.

4.6 Identify Key Factors in Determining the Value of Personalized Medicine

There are currently no consolidated databases of economic evaluations of PM and thus it is difficult to assess the value of PM and what factors determine its value. A consolidated database would enable comparisons across studies and interventions. Furthermore, if this database had information on the key factors that may impact value and adoption, it could provide greater understanding of how to maximize the value of PM. Although the Tufts Cost-Effectiveness Registry is a relevant database, it only includes a specific type of CEAs – cost-utility analyses – and it does not include the range of factors that could determine PM's value and adoption.

Factors that could determine the value and adoption of PM include clinical and disease characteristics, test characteristics, and measures of ethical, legal, and social implications. For example:

- Clinical characteristics of relevant disease
- Health and economic burden of relevant disease
- Analytical validity, clinical validity and clinical utility of test
- Availability of testing
- Regulatory status
- Insurance coverage and reimbursement
- Patient and provider preferences
- Ethical concerns
- Impact on family members
- Stakeholder and advocacy support
- Technology assessment and clinical guidelines
- Impact on diverse populations

Such a database could then be used to examine topics such as:

- Assess the overall cost-effectiveness of PM technologies relative to other interventions.
- Compare and contrast the cost-effectiveness of PM applications, including analyses stratified by the specific test or specific disease.
- Identify factors that influence the cost-effectiveness of PM.
- Identify opportunities to fill gaps and improve the quality of the PM evidence base.
- Evaluate the policy implications of the findings, including which PM technologies get adopted and used.

4.7 Use Real World Perspectives in Economic Analyses

The cost-effectiveness of a PM technology as *actually implemented* may vary from its cost-effectiveness under *ideal* circumstances – and both results can be important depending on the research question. There are several approaches to incorporating "real world" perspectives in economic analyses that could be more widely used. Although a major challenge has been the lack of real world data for inclusion in CEAs, with the advent of more patient centered outcomes research, it is likely that observational data will become more available and of higher quality.

One approach is to use data inputs derived from actual implementation in addition to or in lieu of data derived from controlled conditions. Guidelines recommend that CEAs use the "best" data available as inputs into CEA (Gold et al. 1996). However, there has been limited discussion of whether the "best" data for PM analyses should be derived from *controlled and standardized conditions* – such as randomized clinical trials, high volume central labs, and academic medical centers – versus *real world data* – observational studies, low-volume community labs, and community practices – and the impact on the results. For example, as noted earlier, CEAs of HER2 testing have typically assumed that testing will be offered to the relevant population, that it will be accurate, and that treatment decisions will follow test results – all assumptions that have been questioned when examined using observational data from community settings.

Another approach is to include a base case scenario that reflects actual practice in addition to or in lieu of the "ideal" base case scenario. The choice of comparators in any CEA is a critical determinant of the CEA results. In the case of PM, it can make a large difference in the results if the new PM technology is compared to (1) practice as recommended by guidelines *or* (2) what patients and providers actually do. When there are discrepancies between the two, the choice

of comparator will have a large impact, e.g., a test to target statins where actual rates of taking statins as recommended is much lower than recommended by guidelines.

Lastly, analysts can use extensive sensitivity analyses to examine the implications of how a technology is being implemented or will be implemented. These analyses can examine both changing the structure of the model (e.g., comparing the PM technology to both guidelines and to actual practice) and data inputs (e.g., assessing results using data inputs derived from ideal conditions and from actual practice).

4.8 Consider Patient Heterogeneity in Economic Analyses

Patient centered outcomes research (PCOR) and comparative effectiveness research (CER) have been criticized because they typically focus on characterizing a therapy's average net benefits (Garber and Tunis 2009; Philipson and Sun 2011). There is growing evidence that average net benefits may differ substantially from the net benefits obtained by individuals because of patient heterogeneity (variability). As a result, conventional PCOR/CER analysis may produce misleading results and faulty conclusions about a therapy's value. CEA and other economic evaluations are not part of PCORI's mandate, but the issues in assessing heterogeneity in CER/PCOR and CEA share some similar characteristics.

While there is a growing body of literature examining the impact of patient heterogeneity on treatment effects for CER/PCOR, there is less work on the impact of patient heterogeneity on economic evaluations. A conceptual framework developed by Sculpher demonstrates how calculation of average cost-effectiveness can mask important variations among individual patients (Sculpher 2008). Case studies support this assertion. For example, preliminary findings looking at schizophrenia treatment found that the usual population subgroupings (e.g., age, gender, race) did not explain treatment effect variation and that coverage decisions based on average results are suboptimal (National Pharmaceutical Council 2011). Sculpher notes that the use of subgroup analysis in cost-effectiveness analysis raises a number of methodological questions, including a need to define the possible sources of heterogeneity that exist, to assess how heterogeneity in model parameters should be estimated and how uncertainty should be appropriately quantified, and the appropriateness in terms of equity of using all or some of the subgroup analyses as a basis of decision making.

These issues are specifically relevant to PM because of its focus on individual heterogeneity, particularly in terms of genetic characteristics. Thus methods

developed for analyzing PM can both benefit from and contribute to the development of methods for appropriately assessing heterogeneity in CEA and CER/PCOR. This is particularly true for a broad definition of PM that encompasses populations defined by various personal, clinical, and environmental characteristics such as race, ethnicity, gender, and location.

PM further complicates CEA subgroup analysis because patient heterogeneity includes an array of sociodemographic characteristics and patient preferences, and their interaction with heterogeneity in tumor biology. Consider the case of GEP for breast cancer. Onco*type*DX™, the most commonly used GEP test in the USA, is currently indicated only for early-stage estrogen receptor-positive, lymph-node-negative breast cancers. Yet the aggressive form of breast cancer – triple negative breast tumors (estrogen receptor negative, progesterone-receptor negative, and HER2 negative) – are more common in young women in their 20s and 30s and black women. Thus, in conducting a CEA of GEP, generating subgroup data for black women, for example, may not be relevant for two reasons. First, fewer blacks than whites express estrogen receptor positive tumors, so a CEA focusing on blacks would most likely be inconclusive. Second, even among black women who are estrogen receptor positive, our review of the EGAPP Working Group report on GEP tests (Marchionni et al. 2008) suggest that the clinical utility of GEP had not been validated across diverse populations (Odierna et al. 2011). Thus, for blacks, there is more uncertainty in the use of inputs used in evaluating the cost-effectiveness of Onco*type*DX™.

Efforts to encompass a diverse population perspective in PM CEA could be improved in the following ways:

Increasing Representation: Diversity in PM studies includes both patient level diversity and tumor biology diversity. Recruiting diverse populations with different biomarker expression in clinical trials would reduce the uncertainty of CEA subgroup estimates and make subgroup analysis more meaningful. As a first step, EGAPP could be more explicit in recommending that clinical trials of PM report race/ethnicity of its study participants, so that the assessment of effectiveness across diverse populations is transparent.

Ensuring Collection: Utilization and outcome estimates for subgroups as inputs for PM CEA suffer the same limitations as most other CEAs – self-reported race/ethnicity and self-reported income have a high percentage of missing values in these datasets. An Institute of Medicine (IOM) report (2009) estimates that up to 70% of race/ethnicity are missing in claims data. Imputed data on race/ethnicity offer a proxy measure for analysis, but encouraging health plans to collect self-reported data would facilitate accuracy of inferences from subgroup analysis for PM CEA and CEA in general.

4.9 Prepare for Upcoming Challenges of Assessing Value of Emerging Technologies

The rapid technological advances in PM threaten to outpace our ability to use these interventions effectively in clinical practice and to address the associated health policy issues. This is particularly true for the much awaited "affordable genome" – the ability to sequence an individual's or a tumor's entire genome quickly and inexpensively. Whole genome sequencing is now being offered in clinical care and is expected to become more widely used in the near future, particularly in cancer. Whether whole genome sequencing can achieve its potential to improve patient outcomes will depend on how patients and providers value the information provided, whether whole genome sequencing will be covered by payers and recommended in guidelines, and whether the economic value to the health care delivery system outweighs the costs. Yet, assessing the value of whole genome sequencing is complex because it provides not just one test result but a multitude of results that range from clinically actionable findings (treatable or preventable), to not directly actionable findings (with unclear treatment implications), to findings of unknown significance (Figure 5). There are also concerns that some information could be harmful if there are no available or acceptable treatments or if the information leads to confusion or unwarranted health care.

Thus, an important research need is to develop a framework to conceptualize, identify, and define needed data to assess the value of whole genome sequencing. It would be infeasible and unhelpful to simply propose a CEA of whole genome sequencing versus no whole genome sequencing in its entirety. Such an analysis would be overwhelming in its complexity: there are too many possible research questions, pathways, and outcomes because a whole genome sequencing report

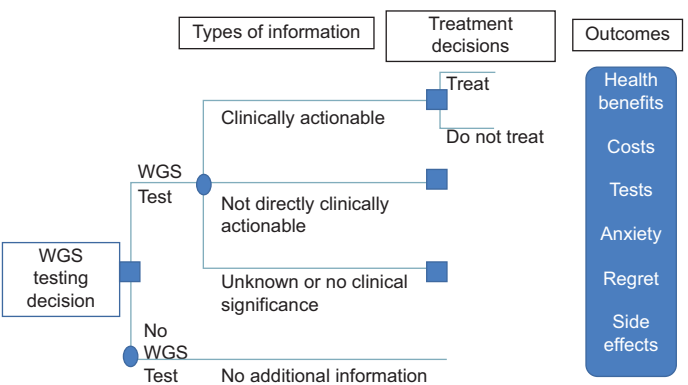

Figure 5 WGS Testing Process and Outcomes.

provides information about multiple variants, each of which has its own probabilities, resulting treatment choices, and downstream consequences to consider, many of which have not yet been defined. Furthermore, a comprehensive analysis of whole genome sequencing needs to consider non-monetary risks and benefits that may not be adequately captured by standard CEA methods.

4.10 Incorporate Behavioral Economics into Value Assessments

Behavioral economics is the combined discipline of psychology and economics that investigates what happens in markets in which agents display human limitations and complications. Seminal work was conducted by Tversky and Kahneman in the 1970s, and Kahneman received the Nobel Prize in Economics for his work in 2002 – the first psychologist to do so. (Tversky and Kahneman 1974; Kahneman and Tversky 1979).

Only recently has behavioral economics been applied to health care (Rice 2013). PubMed did not have a MeSH term for behavioral economics until 2012. There are many examples of where individuals make health decisions that are contrary to rational economic theory, such as failing to enroll in health insurance to which they are entitled or engaging in harmful health behaviors. Behavioral economics offers a means by which to explain and influence such behaviors. One well-known example is the framing effect, an example of cognitive bias, in which people react differently to a particular choice depending on whether it is presented as a loss or as a gain, e.g., individuals perceive "this operation has an 80% survival rate" differently than "you have a 20% chance of dying." Studies applying behavioral economics have looked at issues such as how to incentivize healthier choices and using default choices and "nudges" to improve decision-making (e.g., Halpern et al. 2007; Volpp et al. 2011).

To our knowledge, behavioral economics has not been directly applied to the analysis of decision-making and value for PM. There have been studies of the behavioral impact of genetic testing, such as the Scripps Health study on direct to consumer genetic testing that found that participants who received test results did not suffer test-related distress, but that testing also had little impact on positive behavior changes (Bloss et al. 2011). However, such studies do not appear to have directly used behavioral economic theories.

Behavioral economics could offer rich possibilities for better understanding of the implications of PM and how to most appropriately frame related decision-making at both the individual and collective levels. For example, there are many unanswered questions about how individuals perceive genetic test information

and whether cognitive biases are the same in this type of decision-making as with other decisions or whether "genetic exceptionalism" applies. Studies of PM could also benefit from drawing on the rich literature on cognitive biases in diagnoses, given that most PM involves diagnostic testing. For example, it is well-established that patients and physicians overweight small probabilities and underweight large probabilities, in accordance with prospect theory as developed by Kahneman and Tversky (Elstein and Schwarz 2002). This "compression" of the probability scale explains why the difference between 99 and 100% is psychologically much greater than the difference between, say, 60 and 61%. These issues are particularly important as we consider the implications of whole genome sequencing, because one of the key unanswered questions is how to deal with the dilemma that arises when patients say that they want to know all of the information from sequencing – even if it is useless and may be harmful – given that most patients do not understand the probabilities and tradeoffs involved.

Our work with Reed Johnson at RTI International illustrates how economic analyses could benefit from the application of behavioral analyses (Johnson et al. 2011). This study assessed the consequences of estimating willingness-to-pay assuming a linear opportunity–cost specification when actual preferences may be inconsistent with that specification. Willingness-to-pay estimates derived from stated choice studies generally assume that the marginal utility of income is constant. We analyzed the results of five studies that allow direct tests of this assumption. Tests indicated that marginal utility often violates theoretical expectations. We suggested that this result is an artifact of a cognitive heuristic that recodes cost levels from a numerical scale to qualitative categories. Instead of evaluating nominal costs in the context of a budget constraint, subjects may recode costs into categories such as "low," "medium," and "high," and choose as if the differences between categories were equal. This simplifies the choice task but undermines the validity of willingness-to-pay estimates as welfare measures. Recoding may be a common heuristic in healthcare applications when insurance coverage distorts subjects' perception of the nominal costs presented. Recoding may also distort estimates of marginal rates of substitution for other attributes with numeric levels. Thus, there is an important role in using behavioral economics to better understand how cost is perceived.

In conclusion, PM offers many opportunities as well as challenges. A particularly important challenge is defining and measuring its value. We have outlined six areas for future research, ranging from applying existing methods to PM to breaking new ground through development of new conceptual frameworks and methods. By moving ahead on this research agenda, economists can help ensure that these new technologies have a favorable impact on clinical care and health policy.

References

Armstrong, K. (2012) "Can Genomics Bend the Cost Curve?" Journal of The American Medical Association, 307(10):1031–1032.

Beaulieu, M., S. de Denus and J. Lachaine (2010) "Systematic Review of Pharmacoeconomic Studies of Pharmacogenomic Tests," Pharmacogenomics, 11(11):1573–1590.

Bloss, C. S., N. J. Schork and E. J. Topol (2011) "Effect of Direct-to-Consumer Genomewide Profiling to Assess Disease Risk," New England Journal of Medicine, 364(6):524–534.

Bridges, J. F., A. B. Hauber, D. Marshall, A. Lloyd, L. A. Prosser, D. A. Regier, F. R. Johnson and J. Mauskopf (2011) "Conjoint Analysis Applications in Health–A Checklist: A Report of The ISPOR Good Research Practices for Conjoint Analysis Task Force," Value Health, 14(4):403–413.

Carlson, J. J., N. B. Henrikson, D. L. Veenstra and S. D. Ramsey (2005) "Economic Analyses of Human Genetics Services: A Systematic Review," Genetics in Medicine, 7(8):519–523.

Davis, J. C., L. Furstenthal, A. A. Desai, T. Norris, S. Sutaria, E. Fleming and P. Ma (2009) "The Microeconomics of Personalized Medicine: Today's Challenge and Tomorrow's Promise," Nature Reviews Drug Discovery, 8(4):279–286.

Deverka, P. A., J. Vernon and H. L. McLeod (2010) "Economic Opportunities and Challenges for Pharmacogenomics," Annual Review of Pharmacology Toxicology, 50:423–437.

Djalalov, S., Z. Musa, M. Mendelson, K. Siminovitch and J. Hoch (2011) "A Review of Economic Evaluations of Genetic Testing Services and Interventions (2004–2009)," Genetics in Medicine, 13(2):89–94.

Elkin, E. B., D. A. Marshall, N. A. Kulin, I. L. Ferrusi, M. J. Hassett, U. Ladabaum and K. A. Phillips (2011) "Economic Evaluation of Targeted Cancer Interventions: Critical Review and Recommendations," Genetics in Medicine, 13(10):853–860.

Elstein, A. S. and A. Schwarz (2002) "Clinical Problem Solving and Diagnostic Decision Making: Selective Review of the Cognitive Literature," British Medical Journal, 324(7339):729–732.

Garber, A. M. and S. R. Tunis (2009) "Does Comparative-Effectiveness Research Threaten Personalized Medicine?" New England Journal of Medicine, 360(19):1925–1927.

Giacomini, M., F. Miller and B. J. O'Brien (2003) "Economic Considerations for Health Insurance Coverage of Emerging Genetic Tests," Community Genetics, 6(2):61–73.

Gold, M. E., L. B. Russell, J. E. Siegel and M. E. Weinstein, Eds. (1996) *Cost-Effectiveness in Health and Medicine*. New York, Oxford University Press.

Greenberg, D. and P. J. Neumann (2011) "Does Adjusting for Health-Related Quality of Life Matter in Economic Evaluations of Cancer-Related Interventions?" Expert Review of Pharmacoeconomics & Outcomes Research, 11(1):113–119.

Guo, J. J., S. Pandey, J. Doyle, B. Bian, Y. Lis and D. W. Raisch (2010) "A Review of Quantitative Risk-Benefit Methodologies for Assessing Drug Safety and Efficacy-Report of the ISPOR Risk-Benefit Management Working Group," Value Health, 13(5):657–666.

Haas, J. S., S. Y. Liang, M. J. Hassett, S. Shiboski, E. B. Elkin and K. A. Phillips (2011) "Gene Expression Profile Testing for Breast Cancer and the Use of Chemotherapy, Serious Adverse Effects, and Costs of Care," Breast Cancer Res Treat, 130(2):619–626.

Hall, P., C. McCabe, R. Stein and D. Cameron (2012) "Economic Evaluation of Genomic Test-Directed Chemotherapy for Early-Stage Lymph Node-Positive Breast Cancer," Journal of the National Cancer Institute, 104:56–66.

Halpern, S., P. Ubel and D. Asch (2007) "Harnessing the Power of Default Options to Improve Health Care," New England Journal of Medicine, 357(13):1340–1344.

Institute of Medicine (2009) *Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement*. Washington, DC: The National Academies Press.

Johnson, F. R., A. F. Mohamed, S. Ozdemir, D. A. Marshall and K. A. Phillips (2011) "How Does Cost Matter in Health-Care Discrete-Choice Experiments?" Health Econ, 20(3):323–330.

Kahneman, D. and A. Tversky (1979) "Prospect Theory: An Analysis of Decisions Under Risk," Econometrica, 47(2):263–291.

Khoury, M. J., R. J. Coates, M. L. Fennell, R. E. Glasgow, M. T. Scheuner, S. D. Schully, M. S. Williams and S. B. Clauser (2012a) "Multilevel Research and the Challenges of Implementing Genomic Medicine," Journal Of The National Cancer Institute Monograph, 2012(44):112–120.

Khoury, M. J., M. L. Gwinn, R. E. Glasgow and B. S. Kramer (2012b) "A Population Approach to Precision Medicine," American Journal of Preventive Medicine, 42(6):639–645.

Knight, S. (2011) *Value of Genetic Testing for Heritable Disease in a Population-Based Sample of Older Adults*. Seattle, WA: AcademyHealth.

Kreft, I. G. G. (1995) "Hierarchical Linear Models: Problems and Perspectives," Journal of Education and Behavioral Statistics, 20(2):109–113.

Kuppermann, M., G. Wang, S. Wong, A. Blanco, P. Conrad, S. Nakagawa, J. Terdiman, U. Ladabaum (2012) "Preferences for Outcomes Associated with Decisions to Undergo or Forego Genetic Testing for Lynch Syndrome," Cancer, 119(1):215–225.

Ladabaum, U., G. Wang, J. Terdiman, A. Blanco, M. Kuppermann, C. R. Boland, J. Ford, E. Elkin and K. A. Phillips (2011) "Strategies to Identify the Lynch Syndrome Among Patients with Colorectal Cancer: A Cost-Effectiveness Analysis," Annals of Internal Medicine, 155(2):69–79.

Liang, S.-Y., K. A. Phillips, G. Wang, C. Keohane, J. Armstrong, W. M. Morris and J. S. Haas (2011) "Tradeoffs of Using Administrative Claims and Medical Records to Identify the Use of Personalized Medicine for Patients with Breast Cancer," Medical Care, 49(6):e1–e8. DOI: 10.1097/MLR.1090b1013e318207e318287e.

Marchionni, L., R. F. Wilson, S. S. Marinopoulos, A. C. Wolff, G. Parmigiani, E. B. Bass and S. N. Goodman (2008) *Impact of Gene Expression Profiling Tests on Breast Cancer Outcomes*. Baltimore, MD: Johns Hopkins University Evidence-based Practice Center.

Marshall, D., N. Kulin and I. L. Ferrusi (2012) *Assessing the Value of Personalized Medicine (PM) in Practice - Utilization, Economics and Preferences*. ISPOR 17th Annual International Meeting, Washington, DC.

Marshall, D. A., F. R. Johnson, N. A. Kulin, S. Ozdemir, J. M. Walsh, J. K. Marshall, S. Van Bebber and K. A. Phillips (2009) "How do Physician Assessments of Patient Preferences for Colorectal Cancer Screening Tests Differ from Actual Preferences? A Comparison in Canada and the United States Using a Stated-Choice Survey," Health Econ, 18(12):1420–1439.

Mauskopf, J. A., S. D. Sullivan, L. Annemans, J. Caro, C. D. Mullins, M. Nuijten, E. Orlewska, J. Watkins and P. Trueman (2007) "Principles of Good Practice for Budget Impact Analysis: Report of the ISPOR Task Force on Good Research Practices – Budget Impact Analysis," Value in health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research, 10(5):336–347.

Myers, E., A. J. McBroom, L. Shen, R. E. Posey, R. Gray and G. D. Sanders (2012) *Value-of-information Analysis for Patient-centered Outcomes Research Prioritization*. Durham, NC: Duke University Evidence-based Practice Center.

National Research Council (2011) *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease*. Washington, DC: The National Academies Press.

National Pharmaceutical Council (2011) "NPC Outlines Heterogeneity Issues at AHRQ Methods Symposium," E.V.I.dently Today.

Neumann, P. J. (2005) *Using Cost-Effectiveness Analysis to Improve Health Care: Opportunities and Barriers*. Oxford: Oxford University Press.

Neumann, P. J., J. T. Cohen, J. K. Hammitt, T. W. Concannon, H. R. Auerbach, C. Fang and D. M. Kent (2012) "Willingness-to-Pay for Predictive Tests with No Immediate Treatment Implications: A Survey of US Residents," Health Econ, 21(3):238–251.

Odierna, D., A. Afable-Munsuz, O. Ikediobi, M. Beattie, S. Knight, M. Ko, A. Wilson and N. Ponce (2011) "Early Developments in Gene-Expression Profiling of Breast Tumors: Potential for Increasing Black-White Patient Disparities in Breast Cancer Outcomes?" Personalized Medicine, 8(6):669–679.

Philipson, T. J. and E. Sun (2011) *Blue Pill or Red Pill: The Limits of Comparative Effectiveness Research*. Project FDA, Manhattan Institute.

Phillips, K. A. (2008) "Closing the Evidence Gap in the Use of Emerging Testing Technologies in Clinical Practice," Journal of The American Medical Association, 300(21):2542–2544.

Phillips, K. A., M. J. Ackerman, J. Sakowski and C. I. Berul (2005) "Cost-Effectiveness Analysis of Genetic Testing for Familial Long QT Syndrome in Symptomatic Index Cases," Heart Rhythm, 2(12):1294–1300.

Phillips, K. A., F. R. Johnson and T. Maddala (2002) "Measuring What People Value: A Comparison of 'Attitude' and 'Preference' Surveys," Health Services Research, 37(6):1659–1679.

Phillips, K. A., D. A. Marshall, J. S. Haas, E. B. Elkin, S. Y. Liang, M. J. Hassett, I. Ferrusi, J. E. Brock and S. L. Van Bebber (2009) "Clinical Practice Patterns and Cost Effectiveness of Human Epidermal Growth Receptor 2 Testing Strategies in Breast Cancer Patients," Cancer, 115(22):5166–5174.

Phillips, K. A., S. Van Bebber and A. M. Issa (2006) "Diagnostics and Biomarker Development: Priming the Pipeline," Nature Reviews Drug Discovery, 5(6):463–469.

Phillips, K. A. and S. L. Van Bebber (2004) "A Systematic Review of Cost-Effectiveness Analyses of Pharmacogenomic Interventions," Pharmacogenomics, 5(8):1139–1149.

Phillips, K. A. and S. L. Van Bebber (2005) "Measuring the Value of Pharmacogenomics," Nature Reviews Drug Discovery, 4(6):500–509.

Phillips, K. A., D. Veenstra, S. L. Van Bebber and J. Sakowski (2003) "An Introduction to Cost-Effectiveness and Cost-Benefit Analysis of Pharmacogenomics," Pharmacogenomics, 4(3):231–239.

Phillips, K. A., D. L. Veenstra, E. Oren, J. K. Lee and W. Sadee (2001) "Potential Role of Pharmacogenomics in Reducing Adverse Drug Reactions: A Systematic Review," Journal of the American Medical Association, 286(18):2270–2279.

Phillips, K. A., D. L. Veenstra and W. Sadee (2000) "Implications of the Genetics Revolution for Health Services Research: Pharmacogenomics and Improvements in Drug Therapy," Health Services Research, 35(5 (Part 3)):128–140.

Polsky, D. (2005) "Does Willingness to Pay Per Quality-Adjusted Life Year Bring Us Closer to a Useful Decision Rule for Cost-Effectiveness Analysis?" Medical Decision Making, 25(6):605–606.

Ponce, N. A., J. Tsui, S. Knight, A. Aafable-Munsuz, U. Ladabaum, R. A. Hiatt and J. S. Haas (2012) "Disparities in Cancer Screening in Individuals with a Family History of Breast or Colorectal Cancers," Cancer, 118:1656–1663.

Raman, G., B. Wallace, K. Patel, J. Lau and T. Trikalinos. (2011) *Technology Assessment: Update on Horizon Scans of Genetic Tests Currently Available for Clinical Use in Cancers*. Medford, MA: Tufts University Evidence-based Practice Center.

Rice, T. (2013) "The Behavioral Economics of Health and Health Care," Annual Review of Public Health, 34, PMID:23297657.

Roth, J. A., L. P. Garrison Jr, W. Burke, S. D. Ramsey, R. Carlson and D. L. Veenstra (2011) "Stakeholder Perspectives on a Risk-Benefit Framework for Genetic Testing," Public Health Genomics, 14(2):59–67.

Sculpher, M. (2008) "Subgroups and Heterogeneity in Cost-Effectiveness Analysis," Pharmacoeconomics, 26(9):799–806.

Trosman, J. R., S. L. Van Bebber and K. A. Phillips (2010) "Coverage Policy Development for Personalized Medicine: Private Payer Perspectives on Developing Policy for the 21-Gene Assay," Journal of Oncology Practice, 6(5):238–242.

Trosman, J. R., S. L. Van Bebber and K. A. Phillips (2011) "Health Technology Assessment and Private Payers' Coverage of Personalized Medicine," Journal of Oncology Practice, **7**(3S):18s–24s.

Tversky, A. and D. Kahneman (1974) "Judgment Under Uncertainty: Heuristics and Biases," Science (4157):1124–1131.

Veenstra, D. L., J. A. Roth, L. P. Garrison, Jr., S. D. Ramsey and W. Burke (2010) "A Formal Risk-Benefit Framework for Genomic Tests: Facilitating the Appropriate Translation of Genomics into Clinical Practice," Genetics in Medicine, 12(11):686–693.

Vegter, S., C. Boersma, M. Rozenbaum, B. Wilffert, G. Navis and M. J. Postma (2008) "Pharmacoeconomic Evaluations of Pharmacogenetic and Genomic Screening Programmes: A Systematic Review on Content and Adherence to Guidelines," Pharmacoeconomics, 26(7):569–587.

Volpp, K., D. Asch, R. Galvin and G. Loewenstein (2011) "Redesigning Employee Health Incentives – Lessons from Behavioral Economics," New England Journal of Medicine, 365:388–390.

Wang, G., M. Kuppermann, B. Kim, K. A. Phillips and U. Ladabaum (2012) "Influence of Patient Preferences on the Cost-Effectiveness of Screening for Lynch Syndrome," American Journal of Managed Care, 18(5):e179–185.

Wang, K., H. Zhang, C. S. Bloss, V. Duvvuri, W. Kaye, N. J. Schork, W. Berrettini and H. Hakonarson (2011) "A Genome-Wide Association Study on Common SNPs and Rare CNVs in Anorexia Nervosa," Molecular Psychiatry, 16(9):949–959.

Wong, W. B., J. J. Carlson, R. Thariani and D. L. Veenstra (2010) "Cost Effectiveness of Pharmacogenomics: A Critical and Systematic Review," Pharmacoeconomics, 28(11):1001–1013.