Debating Clinical Utility

W. Burke\textsuperscript{a}  A.-M. Laberge\textsuperscript{c}  N. Press\textsuperscript{b}

\textsuperscript{a}Department of Bioethics and Humanities, University of Washington, Seattle, Wash., and \textsuperscript{b}Schools of Nursing and Medicine, Oregon Health and Science University, Portland, Oreg., USA; \textsuperscript{c}Service de génétique médicale, CHU Sainte-Justine, Département de Pédiatrie, Université de Montréal, Montréal, Que., Canada

\textbf{Introduction}

The term ‘clinical utility’ was coined by a US task force [1] to describe one of 3 key measures of a genetic test. It was defined as ‘the benefits and risks that accrue from both positive and negative test results’. The other measures were analytic validity, the accuracy with which an assay measures a particular genetic characteristic, and clinical validity, the accuracy with which a genetic characteristic identifies a disease condition or risk. These properties are not independent: a test with poor analytic and/or clinical validity is unlikely to have clinical utility. In this framework, however, analytic and clinical validity are technical properties, while clinical utility addresses a test’s health care value [2–4]. Like other measures of value, it is often contested.

The reasons for disagreement vary. Stakeholders may have different views about benefits and risks and about the importance of social versus health outcomes. They also commonly disagree about the evidence needed to determine whether a test is effective in achieving a specific outcome. Questions may be presented as factual disagreements, when they are actually debates about what information matters or how facts should be interpreted and used in clinical decision-making. Defining the different issues at stake is therefore an important element of policy-making. Key issues include evidence standards for test use, and in particular, the circumstances under which prospective controlled data should be required, as well as evidence on feasibility, cost and equitable delivery of testing; the goals of population-based screening programs, and in particular, the role of social outcomes in evaluating test value; and the appropriate uses and funding of tests that inform non-medical actions. Addressing each of these issues requires attention to stakeholder values and methods for effective deliberation that incorporate consumer as well as health professional perspectives.

\textbf{Key Words}
Clinical utility • Evidence-based practice • Genetic testing

\textbf{Abstract}

The clinical utility of genetic tests is determined by the outcomes following test use. Like other measures of value, it is often contested. Stakeholders may have different views about benefits and risks and about the importance of social versus health outcomes. They also commonly disagree about the evidence needed to determine whether a test is effective in achieving a specific outcome. Questions may be presented as factual disagreements, when they are actually debates about what information matters or how facts should be interpreted and used in clinical decision-making. Defining the different issues at stake is therefore an important element of policy-making. Key issues include evidence standards for test use, and in particular, the circumstances under which prospective controlled data should be required, as well as evidence on feasibility, cost and equitable delivery of testing; the goals of population-based screening programs, and in particular, the role of social outcomes in evaluating test value; and the appropriate uses and funding of tests that inform non-medical actions. Addressing each of these issues requires attention to stakeholder values and methods for effective deliberation that incorporate consumer as well as health professional perspectives.
testing are all influenced by judgments about clinical utility. Underlying value judgments, and related priority-setting decisions, may not always be acknowledged. Instead, questions may be presented as factual disagreements, when they are actually debates about how facts should be interpreted or used in clinical decision-making. Defining the different issues at stake is therefore an important, although often overlooked, element of policymaking and may help to identify barriers to consensus and the strategies needed to resolve them.

**Evidence Thresholds for Genetic Test Use**

New genetic tests are a product of scientific research. Yet the specific evidence needed to justify a test’s clinical use is a frequent source of disagreement. As an example, Blue Cross Health Tec Assessment [5] and the American Society of Clinical Oncology [6] have endorsed gene expression profiling as a means to characterize breast cancer prognosis and inform chemotherapy decisions. By contrast, the Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group, an evaluation group sponsored by CDC, reviewed the same evidence and found it insufficient to recommend for or against such testing [7]. Similarly, experts have disagreed about whether the available evidence is sufficient to recommend pharmacogenetic testing to guide the use of the anticoagulant warfarin [8–11].

Although these debates typically focus on the findings of specific studies (or the absence of studies), the underlying disagreement is about type of evidence needed to justify test use. A core issue is the degree to which different types of clinical studies provide valid outcome data; a common related question is whether prospective evidence on test outcomes should be required prior to test use. Both of these questions relate to the clinical evidence used to establish the test’s potential to achieve its intended purpose. A number of other measures, related to its acceptability, cost and feasibility, are also important in evaluating a test’s clinical utility (table 1) [12–14].

**Evaluating a Test’s Potential to Achieve the Intended Clinical Outcome**

The standard used in the evaluation of new drugs – randomized controlled trials – has not been applied to medical tests. Instead, plausible observational data have traditionally been viewed as sufficient to justify a new test. For example, a new method for measuring blood chemistry is evaluated by demonstrating that its results are either comparable to or better than a gold standard (thus establishing analytical and/or clinical validity), rather than by evidence that measurement of the analyte in question improves patient outcomes (which would establish clinical utility).

This standard works well when there is an accepted clinical role for the test. However, many genetic tests create new clinical paradigms. Take the warfarin example: Tests for variants in the CYP2C9 and VKORC1 genes can identify individuals with lower dosing requirements and a higher risk of bleeding complications from the anticoagulant warfarin; these variants are estimated to account for 30–50% of the individual variation in drug response. Using tests for CYP2C9 and VKORC1 variants to make decisions about warfarin dosing represents a new way to manage drug therapy. Some estimate it will lead to markedly increased drug safety and reduced health care costs [16, 17], while others caution that the outcome is difficult to predict and could in fact have limited benefit, lead to increased costs, and potentially result in errors in drug prescribing [8]. Much of the data supporting pharmacogenetic testing for warfarin derives from retrospective studies [e.g. 18]. Three small clinical trials have been reported, but these were of variable quality, with short follow-up times, and did not provide evidence for significant outcome benefits [19]. Modeling studies have provided evidence for and against cost-effectiveness, with some variables difficult to estimate accurately because of limited empiric data [16, 20]. The debate is therefore fundamentally about the weight given to presumptive benefits and harms in the face of uncertainty and about the trade-offs between bringing a potentially beneficial innovation to health care early versus waiting for more robust evidence. This is a particularly important question in a context of limited resources.

Clinical practice is replete with innovations that proved less beneficial when tested in randomized trials than they initially appeared in observational studies – hormone replacement therapy is a recent example [21]. However, medical genetics offers important counterexamples. Genetic testing for multiple endocrine neoplasia type 2 (MEN 2), followed by prophylactic thyroidectomy in those found to have the condition, was established as a practice standard based solely on observational data; a 5-year follow-up of treated individuals confirms effective prevention of thyroid cancer [23].

The rarity of MEN 2, and the urgency of preventing medullary thyroid cancer in children at risk, arguably made a randomized trial for prophylactic thyroidectomy both impractical and unethical. But to what extent should
this example inform the introduction of other genetic tests? Important factors in addressing this question include the prevalence of the disorder, the predictive value of the test, and the availability and utility of alternative diagnostic tools or treatment regimens. In the MEN 2 example, penetrance of causative mutations is close to 100%, and surveillance tools were inadequate to identify thyroid cancer effectively at an early stage. In the warfarin example, pharmacogenetic tests explain no more than half of individual variation in drug response [11, 15], and other alternatives to safe dosing are available, including the standard clinical approach of initiating warfarin therapy with low initial doses and regularly monitoring the patient’s response. Testing for warfarin treatment is therefore substantively different from testing for MEN 2, but how the difference should inform evidence requirements is a matter of judgment.

A related issue is the generalizability of clinical evidence. A clinical trial may provide evidence for efficacy of the testing process and its follow-up services. However, the benefit achieved under routine conditions – that is, the effectiveness of the testing process – may be lower, due to factors such as provider preparedness, the availability and convenience of follow-up services, and patient compliance. These points speak to the importance of evidence beyond that provided by clinical research (table 1).

**Considering the Full Array of Contributors to Clinical Utility**

The scope of the evidence needed to provide a convincing justification for test use inevitably varies for different tests and clinical settings [9, 12]. For example, another pharmacogenetic test, to identify people at increased risk for adverse effects from the anti-retroviral abacovir, has been greeted with general enthusiasm [24, 25]. The difference in the acceptance of this test compared to warfarin-related testing likely lies in the high specificity of the test: although fewer than 50% of people

---

**Table 1. Contributors to clinical utility**

<table>
<thead>
<tr>
<th>Component</th>
<th>Definition</th>
<th>Relevance to clinical utility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analytic validity</td>
<td>Accuracy of test assay compared to gold standard measure</td>
<td>Determines whether test meets acceptable analytic standard</td>
</tr>
<tr>
<td>Clinical validity</td>
<td>Association of genetic characteristic with specified health condition or risk, and sensitivity, specificity and predictive value in the population groups and clinical setting proposed for testing</td>
<td>Determines potential clinical uses of test</td>
</tr>
<tr>
<td>Test setting and purpose</td>
<td>Clinical and/or demographic description of group(s) to be offered testing, health condition tested for, and specific clinical goal of testing, including any associated services</td>
<td>Determines outcomes sought from testing</td>
</tr>
<tr>
<td>Societal legitimacy</td>
<td>Consistency of testing and associated services with ethical principles, values, norms, mores, laws and regulations</td>
<td>Determines whether test is compatible with societal expectations</td>
</tr>
<tr>
<td>Efficacy and effectiveness</td>
<td>The ability of the test and any associated services to achieve the intended health purpose under the most favorable circumstances (efficacy) and under routine conditions (effectiveness)</td>
<td>Determines the potential for the test to achieve the health outcomes sought</td>
</tr>
<tr>
<td>Balance of outcomes</td>
<td>Assessment of negative relative to positive outcomes of testing and associated services for the person tested</td>
<td>Determines whether testing provides a net benefit to the person tested</td>
</tr>
<tr>
<td>Patient and family acceptability</td>
<td>Consistency of testing and associated services with the wishes, desires, and expectations of patients and their families</td>
<td>Determines whether testing is compatible with patient and family preferences</td>
</tr>
<tr>
<td>Economic measures</td>
<td>The ability of the test and associated services to lower the costs of care without diminishing benefits and/or provide an appropriate health value for the investment of resources</td>
<td>Determines benefits provided by testing relative to investment of resources</td>
</tr>
<tr>
<td>Equity</td>
<td>Access to test and associated services among patients who can benefit</td>
<td>Determines whether testing is compatible with equitable health care delivery</td>
</tr>
</tbody>
</table>

Adapted from Burke et al., 2007 [12]; Burke and Zimmern, 2007 [13]; and Zimmern and Kroese, 2007 [14].
with the risk genotype, HLA-B*5701, will experience adverse events, people without this genotype appear to have no risk [24]. As a result, the pharmacogenetic test offers the physician clinically useful information about patients at risk; given alternative therapies for these individuals, the test has clear clinical utility. Even here, however, contextual factors need to be taken into account. A recent study predicted, not surprisingly, that the cost-effectiveness of HLA-B*5701 testing prior to abacavir use would vary widely with the prevalence of the variant, the costs of both the test and alternative treatments, and the relative effectiveness of the alternative treatments [26]. Under some scenarios the test was highly cost-effective while under others it provided little benefit; under some scenarios the use of an alternative drug without testing was preferable. This example points to the importance of defining the clinical context before evaluating clinical utility (table 1), including the population to be tested and the services to be offered after testing, as well as cost, acceptability and other social factors.

Differing judgments about clinical utility illustrate the central role of evidence standards and related questions about the types of evidence needed, and how they contribute to decision-making, in most debates about the use of genetic tests. Relatively few medical innovations have been established through randomized clinical trials; even when a prospective clinical trial provides evidence of benefit, clinicians must make judgments about the relevance of the trial for their patients, who may differ from trial participants in significant ways [27]. Health service context, societal and patient acceptance, and financial considerations are also relevant (table 1). As a result, there is no single ‘right’ answer in these debates. Clarity about the reason for differences – in particular, why observational or other data are persuasive for some observers, while others remain unconvinced without prospective trial data – may help to inform clinicians and patients who must make decisions about test use. Furthermore, evidence is not static: new studies might lead to a re-evaluation of the clinical utility of a particular test. Ultimately, clarity about the value judgments different stakeholders use in judging evidence can promote broader consensus.

Genetic Testing for Population-Based Disease Prevention

Additional questions about value arise when genetics is proposed as a tool for population-based disease prevention. The use of genetics for this purpose is already well established. The identification of newborns who require urgent treatment to prevent death or disability – as in the case of phenylketonuria – represents the most dramatic example. Another routine use of genetics for disease prevention involves the evaluation of family history to detect individuals at increased risk of cancer and other adult onset diseases, in order to enable targeted prevention. However, debates about clinical value occur for both these uses of genetic information, centering on the implications of a test’s predictive value and the effectiveness of interventions to reduce risk. Increasingly, discussions about genetics and disease prevention also raise questions about the appropriate scope of genetic risk assessment.

Newborn Screening

The development of tandem mass spectrometry has allowed a large increase in the number of conditions tested for in newborn screening, and DNA-based testing offers the potential for further expansion in the future [28]. This growing technological capacity has aroused vigorous debate about the threshold for introducing new tests and, ultimately, about the purpose of this population screening program. As Grosse et al. have pointed out [29], newborn screening was initially instituted to address a public health emergency – the need for rapid institution of diet therapy for infants with phenylketonuria – to prevent mental retardation. Over time, however, the goal of newborn screening has expanded to include detection of infants who do not require immediate treatment, but who will benefit from specialized services – for example, infants with cystic fibrosis. With such expansion comes an increasing number of false-positive findings [30] and the detection of infants with ambiguous test results [31], both adding cost and posing potential harms.

The diagnostic capacity of tandem mass spectrometry also allows for the identification of conditions for which no proven therapy is currently available [32]. In this context, some advocates have proposed that the traditional goal of newborn screening – the improved health of the infants tested – should be expanded to encompass goals related to the family’s quality of life. They note that many parents express a preference for knowing early about an affected child, even if no treatment is available [33]. Early detection of an untreatable genetic disease can also inform reproductive decision-making in future pregnancies [33, 34]. Broad detection of infants with rare genetic diseases is also seen as a way to expedite research [34, 35]. Others argue forcefully against the expansion of newborn screening programs for these purposes [32, 36–38].
The values at stake in this debate include the appropriate uses of a publicly funded screening program [36]; concerns about the lack of explicit informed consent or pre-test counseling in newborn screening programs [38]; potential harms from treatments of unproven value [32, 37]; and concerns about expanding the burden of false-positive test results [30]. These debates are partly about evidence – for example, what evidence is needed to assess the harms of false-positive results – but much more about the values that should inform population screening of newborns. In particular, the debate centers on what concerns or risks justify providing unsought information to parents of healthy infants. The newborn screening example thus illustrates that some contributors to clinical utility – including acceptability of testing from societal and patient perspectives, financial trade-offs, and the balance of positive and negative consequences of testing (table 1) – cannot be assessed without also considering whose views matter and how they should be weighed and incorporated in decision-making.

Detection of Common Disease Risk

An important goal of family history assessment is to identify increased risk for common complex diseases, so that targeted preventive care can be offered. Public campaigns encourage individuals to seek out family history information [e.g. 39], and geneticists have called for increasing clinician education on the use of family history information in disease prevention [40]. Unfortunately, family history is a relatively crude measure for assessing risk for common complex diseases [41].

Recent progress in the identification of gene variants associated with common disease risk [42] points to a new approach to achieving the same goal: personal genomic profiling to identify risk and guide preventive care. Personal genome profiles are already being marketed directly to consumers as a source of health and personal information [43, 44]. Advocates believe that such information could motivate healthy behaviors such as improved diet and exercise or smoking cessation [45–47], and several studies have been launched to seek evidence evaluating the use of such information to improve disease prevention [e.g. 48–50]. Others question the value of this approach [51, 52].

One aspect of the debate focuses on the need for evidence of improved outcomes from genetic testing – a continuation of the debates about outcome data for tests such as warfarin pharmacogenetics and gene expression profiling in breast cancer [53]. In addition, because most gene variants associated with common complex diseases confer very small risks, there is currently uncertainty about the extent to which genomic profiling will provide an effective basis for preventive care [51, 52, 54, 55].

As studies are completed and the scope of benefit is defined, questions about values will arise: How big a prevention effect is sufficient to justify genetic testing? If the main outcome of testing is to suggest a better diet, or other lifestyle improvement, is testing an appropriate use of health care dollars? And is the test still of value if the recipient does not make the lifestyle changes? Consumers may wish to have the option for such genetic testing, and some may argue that they have a right to such information. Resolution of the underlying evidence question – what data are needed to establish the clinical utility of genetic susceptibility testing – will depend on how one views the goals of health care and, in particular, the appropriate role of consumer preference when medical outcomes are uncertain. Costs and associated trade-offs are also a legitimate part of the discussion – in particular, whether expenditures for personal genomics can be justified if they draw resources away from other health expenditures.

Genetic Testing to Inform Non-Medical Decisions

A related question concerns the role of genetic testing in providing information for decisions that are more social than medical. The use of genetic testing for reproductive decision-making provides an interesting precedent for this discussion.

Prenatal genetic testing was introduced at approximately the same time as newborn screening [56]. Carrier tests for a number of genetic diseases soon followed. Although these tests are often discussed in conventional medical terms – e.g. a prenatal test may be described as ‘indicated’ when a pregnant woman is known to be at risk to have a child with a genetic disease – their purpose is different from most medical tests. Rather than informing the health care of the individual tested, carrier and prenatal genetic tests inform parents about the risks of having a child with a genetic disease. In most clinical settings they are offered to enable parents to consider pregnancy termination if a serious genetic disease is identified in the fetus, or to help parents prepare for a child with special needs.

Both societal and personal values inform this testing process. In some countries, the introduction of prenatal diagnosis and access to pregnancy termination have been tied explicitly to societal concerns about the burdens of a
genetic disease – for example, in screening programs for \(\beta\)-thalassemia in Cyprus and Iran [57, 58]. In countries where this service is available, health care providers generally articulate a strong commitment to pre-test counseling, to ensure that testing is voluntary and in keeping with parental preferences.

 Debates around reproductive genetics have focused on the moral implications of pregnancy termination. Many disability advocates have questioned the use of prenatal diagnosis to prevent births of children with Down syndrome, for example [59]. With the introduction of pre-implantation genetic diagnosis, other uses of reproductive genetics – e.g. testing to detect adult onset conditions, or to determine whether the embryo can serve as a bone marrow donor for an ailing sibling – are also controversial [60, 61]. These debates are to be expected, given the nature and purpose of reproductive genetic testing; societal legitimacy (table 1) is a factor in determining what prenatal tests can be offered.

 However, genetic testing can inform personal decision-making in a variety of other ways, raising questions that are analogous to – and ultimately part of – the debate about personal genomics. For example, learning that a child has X-linked retinitis pigmentosa may be extremely important for educational and career planning because the child can be expected to be legally blind by early adulthood [62]. Although the diagnosis provides a clinical prognosis, no specific therapy is currently available to reduce or ameliorate vision loss; as a result, the social uses of the information are more important than the clinical uses.

 The clinical utility of genetic testing for retinitis pigmentosa is unlikely to be questioned because of the high predictive value of a positive test and the specific preparatory actions that can be taken by parents and affected persons. Less predictive genetic tests offer information that individuals may find similarly useful for life planning, but these tests are likely to be more controversial. As an example, APOE 4 testing can identify individuals at increased risk of Alzheimer disease. A small study found that those with positive test results were more likely to purchase long-term care insurance [63], and preparing family members was viewed as an important value of testing [64]. Yet several expert panels have recommended against such testing, on the grounds that the predictive value of testing is limited and the risk information could be stigmatizing and emotionally upsetting [65–67]. A recent study indicating lack of short-term psychological stress after APOE 4 testing [68] will not necessarily reduce these concerns, given that the participants in this study were unlikely to be broadly representative of the population [69]. These differences of opinion reflect different estimates of the benefits and risks associated with probabilistic information and perhaps also reflect different stakeholders’ views about the goals of health care and appropriate uses of health care resources. Over the next decade, genomic research will offer many additional tests to fuel this debate.

 Benefits of Defining the Issues

 Lack of evidence has been identified as a major impediment to the translation of genomic knowledge into beneficial medical interventions [49, 53, 70]. However, the task of defining what is adequate evidence may, in fact, be at the heart of many disputes and will need to be considered in developing consensus on clinical utility.

 Perhaps the first issue to be addressed is whether ‘clinical utility’ should be considered relevant only in health care settings. A test that provides information of interest to consumers but is not medically actionable, like the APOE 4 test, might have a poor claim on health care resources [71], yet might still represent an appropriate consumer product. If so, consumer safety would become a central policy concern, with a need to define the potential harms of testing, the regulatory models for pre-market test review, and the standards for the marketing of products [43]. As debates about personal genomics already demonstrate, defining the line between consumer products and health care tests will also be difficult.

 For tests used in health care, evidence standards will need to be based on what physicians, patients, and health care funders find convincing in establishing a benefit. For example, will a genetic risk assessment that is believed to motivate a change in patient behavior, rather than changes in physician testing or prescribing regimens, be considered medically actionable and thus worthy of a claim on health care dollars? The threshold defined by clinicians in practice may or may not conform to the rigorous standards proposed by groups such as EGAPP [72] – and patients may view the threshold differently than clinicians.

 Some will argue that clinicians in practice are ill equipped to assess the clinical utility of new genetic tests. Most have important deficiencies in their knowledge of genetics and genetic tests [73], and most medical students do not retain the genetics education they received [74, 75]. It would therefore be unrealistic to presume that most clinicians will be able to integrate new genetic tests into
their practice based on their assessment of the evidence. Public health efforts to increase the development of practice guidelines in genetics are underway [72, 76]. There is a need for greater physician engagement in the development and use of guidelines and more systematic efforts to assess the large number of genetic tests likely to emerge from current research [77], with appropriate stakeholder input.

The evidence needed to make a compelling case for testing will undoubtedly vary by both test characteristics and testing purpose [12]. The clinical utility of tests to diagnose rare, highly penetrant conditions will generally be established by small-scale studies that confirm the gene–disease association. On the other hand, tests for genetic susceptibility, intended to be used in population-based screening, are unlikely to be convincing without rigorous assessment of testing outcomes.

Clarification of different stakeholders whose interests are at stake, and their preferences and values, will also be important. In some cases – such as the use of testing to inform medical treatment of symptomatic patients – little controversy will be expected, and a convergence of values can be predicted. However, in other arenas, such as medical testing used for actions outside the medical system (e.g. APOE testing to inform personal decisions such as purchase of long-term care insurance) or population screening for rare conditions with variable phenotype and severity, controversy is to be expected. Stakeholders for these decisions include not only clinicians, patients and health care funders, but also test developers, regulatory agencies and lawmakers. In these latter cases, endless debate without resolution can occur – and clarifying the values that are at stake and how different stakeholders prioritize them may be the only way to move discussion forward to a resolution.

An early challenge in approaching this task is to determine how different stakeholder views can be defined and shared. While there are good reasons to separate the processes of regulatory review, development of professional practice guidelines, and funding decisions – because they are based on different governance – more opportunities are needed to discuss the different values that may be brought to each of these decision-making activities. Perhaps more important, with increasing attention to patient-centered care [78], there is a need to move beyond expert-driven processes, to identify ways for meaningful input from the consumers who are both the intended beneficiaries and ultimate funders of genomic innovation.

Acknowledgement

This project was supported in part by the Center for Genomics and Healthcare Equality (Grant P50 HG003374 from the US National Institutes of Health).

References


71 McGuire AL, Burke W: An unwelcome side effect of direct to consumer personal genome testing: raiding the medical commons. JAMA 2008;300:2669–2671.


77 McGuire AL, Burke W: An unwelcome side effect of direct to consumer personal genome testing: raiding the medical commons. JAMA 2008;300:2669–2671.