

Evidence-based classification of recommendations on use of genomic tests in clinical practice: Dealing with insufficient evidence

Muin J. Khoury, MD, PhD¹, Ralph J. Coates, PhD¹, and James P. Evans, MD, PhD²

Abstract: Numerous genomic tests continue to emerge as potential tools in the diagnosis, treatment, prognosis, and prevention for a wide variety of common human diseases. To date, most of these tests have “insufficient evidence” of clinical validity and utility for their use in clinical practice. Explicit and quantitative tools can be used in the evaluation of direct and indirect evidence on the utility of genomic tests. As suggested in an article in this month’s issue by Veenstra et al., a recommendation matrix can be developed based on the amount of certainty of the evidence and the assessment of the risk-benefit profile. To supplement the current binary (up or down) evidence-based recommendation for use, it is worthwhile to explore all available data to develop a three-tier evidence-based recommendation classification of genomic tests (“use in practice,” “promote informed decision-making,” and “discourage use”). Promoting informed decision making may be a valuable recommendation for tests for which there is sufficient information on analytic and clinical validity and for which the risk/benefit analysis on clinical utility is promising but not definitive. This approach could provide interim guidance for clinical practice, while rigorous outcomes research is conducted to assess the impact of such tests on patients, families, and population health outcomes. *Genet Med* 2010;12(11):680–683.

Key Words: *evidence-based medicine, genomic medicine, public health*

“Decision makers do not have the luxury of waiting for certain evidence. Even though evidence is insufficient, the clinician must still provide advice, patients must make choices, and policymakers must establish policies.”¹

In this issue of the *Genetics in Medicine*, Veenstra et al.² present a formal risk-benefit framework for assessing the health-related utility of genomic tests. Their approach is based on combining methods from the fields of decision science, outcomes research, and health technology assessment. Their framework entails (1) using decision analysis to synthesize data, project incidence of health outcomes, and assess uncertainty; (2) defining health-related utility of genomic tests as improvement in health outcomes as measured by quality-adjusted life-years; and (3) displaying results using a risk-benefit matrix to facilitate the interpretation of findings from these analyses. The matrix leads to a classification of genomic tests based on the risk-benefit profile and the amount of uncertainty. Such a classification could inform decisions about use of genomic tests in

practice. Veenstra et al.² discuss the strengths and limitations of this approach and the crucial need for stakeholder engagement. In this commentary, we put the work by Veenstra et al.² in the context of the evidentiary challenge of translating genomic discoveries into health benefits. We also discuss the issue of “insufficient evidence” facing genomic medicine and promote the implementation of evidence-based triaging of genomic applications for specific intended uses in practice.

THE EVIDENTIARY CHALLENGE OF GENOMIC MEDICINE

The approach of Veenstra et al.² is motivated largely by the ongoing frustration with the lack of an evidentiary basis for the translation of genomic discoveries into clinical practice. We subscribe to the notion that genetic and genomic information is, by and large, similar to other medical- or health-related information, and thus, it should require empirical evidence of clinical utility in practice (net positive health impact minus potential harms).³ The field of genomics faces several challenges as Veenstra et al.² discuss. First, there is a lack of comparative outcomes data for genomic applications due to regulatory and reimbursement policies that do not require such studies⁴ and their inherent costs. In addition, there is a relative ease of market access for genomic tests, including direct to consumer testing, which makes the lack of evidence more problematic.⁵ Finally, there is no consensus on evidentiary requirements for genomic test evaluation. Some stakeholders accept the findings of observational studies or even biological plausibility of potential benefits, whereas others insist on randomized controlled clinical trials.⁶

In the United States, two independent evidentiary groups have attempted to address the issue of how to evaluate genomic tests along with a variety of other health services.^{7–10} The Advisory Committee on Heritable Disorders in Newborns and Children evaluates genomic tests for use in newborn screening panels.¹¹ The Evaluation of Genomic Applications in Practice and Prevention Working Group (EGAPP) provides an evidence-based assessment of genomic tests and other applications that are in transition from research to clinical and public health practice.¹² Since 2005, EGAPP has developed model approaches, commissioned evidence reviews, and made recommendations on four genomic tests^{13–16} and several more are on the way. EGAPP adapted methods of other evidentiary bodies such as the US Preventive Services Task Force (USPSTF), which has evaluated and made recommendations about clinical preventive services in the United States for more than 2 decades.^{1,7,17} Although the overall process has been relatively slow compared with the blistering pace of emerging technologies and its applications, EGAPP has laid an important initial methodological foundation for how to go about evaluating genomic applications in an evidence-based and transparent

From the ¹Office of Public Health Genomics, Centers for Disease Control and Prevention, Atlanta, Georgia; and ²Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina.

Dr. Muin J. Khoury, Office of Public Health Genomics, CDC, 1600 Clifton Road, Atlanta, GA 30333. E-mail: mkhoury@cdc.gov.

Disclaimer: The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Disclosure: The authors declare no conflict of interest.

Published online ahead of print October 22, 2010.

DOI: 10.1097/GIM.0b013e3181f9ad55

fashion. The working group has recently published their methodology and approach to topic selection, developing analytic frameworks with direct (e.g., using data from randomized clinical trials) and indirect evidence (using a causal chain of evidence with all available studies) on clinical utility.¹⁸ They have focused on a detailed assessment of analytic validity, clinical validity, and clinical utility by type of application (diagnostic, risk assessment, etc.) and by intended use for target populations and a detailed rationale for their recommendations. More recently, they have described endpoints of interest in evaluating health-related outcomes such as diagnostic thinking, therapeutic choice, patient outcome impact and familial and societal impacts, all of which are important in evaluating the utility of genomic tests.¹⁹ The working group has faced many of the challenges raised by Veenstra et al.² when deliberating on their methods and outcomes and when conducting evaluations for the first few topics. As we discuss later, the recommendation by Veenstra et al.² for an evidence-based classification of genomic tests is synergistic with the methods paper of EGAPP. Moreover, it could provide a practical foundation for a more rapid implementation of evidence-based triaging of genomic tests.

THE ISSUE OF “INSUFFICIENT EVIDENCE” IN GENOMIC MEDICINE

A major issue we face in the rapidly developing field of genomics is “insufficient evidence.” Out of the first four genetic tests assessed by EGAPP, three returned “insufficient evidence to recommend for or against.”^{13–16} We suspect that many more emerging genomic tests will end up in this category. From September 2009 to July 2010, approximately 200 “new” genomic tests were making their way from the bench to the bedside.²⁰ For most of these tests, informed readers can quickly infer insufficient evidence on their validity and utility for use in routine practice. However, when an evidence-based group spends time and resources to arrive at “insufficient evidence,” this may seem frustrating to clinicians, consumers, and policy makers who still have to make decisions on the basis of insufficient evidence.

The problem of insufficient evidence, of course, is not unique to genomics,¹ again emphasizing that approaching problems inherent to the field of genetics can benefit greatly from how the same issues are dealt with in other medical contexts. Indeed, the well-established USPSTF has found insufficient evidence for a large number of services, even for services for which substantial research has been conducted, e.g., mammography screening for breast cancer between 40 and 50 years of age.²¹ The problem of insufficient evidence is exacerbated by the evidence review findings that for many topics considered by the USPSTF, there is limited research using a comparative effectiveness approach. Current research most often does not permit even moderate certainty about the net benefit of the preventive service. In addition, evidence about the net benefits of preventive services in subgroups defined by age, sex, race, and other factors is likely to remain perpetually uncertain because additional subgroup questions are defined once evidence is obtained.¹ Although the words “genetics” or “genomics” were not mentioned by the USPSTF, classifications of populations by genetic or genomic tests will also define subgroups in which evidence may remain uncertain.

In 2009, the USPSTF noted that clinician stakeholders have commented that recommendations of insufficient evidence are not really recommendations, and some have even characterized a recommendation of insufficient evidence as useless, or even “worse than useless.”¹

INFORMED DECISION MAKING WHEN EVIDENCE IS INSUFFICIENT

As the USPSTF has long recognized, even though evidence may be insufficient, the clinician must still provide advice, patients must make choices, and policymakers must establish policies.¹ In 2004, the USPSTF recommended that clinicians use shared decision making as an appropriate approach for services for which evidence was insufficient, or the balance of benefits and harms was weakly positive or would vary depending on individual values or preferences.²² The Task Force defined shared decision making as a particular process of decision making by the patient and clinician in which the patient (1) understands the risk or seriousness of the disease or condition to be prevented; (2) understands the preventive service, including the risks, benefits, alternatives, and uncertainties; (3) has weighed his or her values regarding the potential benefits and harms associated with the service; and (4) has engaged in decision making at a level at which he or she desires and feels comfortable. This process has the goal of an informed and joint decision. The Task Force on Community Preventive Services has made similar recommendations about decision making in the context of insufficient evidence.²³ They recommended “informed decision-making,” and, working with the USPSTF, they identified shared decision making as a subcategory of the broader informed decision-making approach. According to the Task force, informed decision making occurs when an individual understands the nature of the disease or condition being addressed by the service; understands the clinical service and its likely consequences, including benefits, harms, limitations, alternatives, and uncertainties; has considers his or her preferences as appropriate; has participates in decision making at a personally desirable level; and either makes a decision consistent with his or her preferences and values or elects to defer a decision to a later time.

In 2009, in response to continuing comments from stakeholders about recommendations of insufficient evidence, the USPSTF began to provide additional information to clinicians to help with decision making in the context of insufficient information on the balance of harms and benefits of a service.¹ This information included four “domains” important to decision making: the potential burden of disease that might be prevented by an effective service, the potential harms from such a service, the costs—including opportunity costs—of widespread use of the service, and a description of current practice.¹ This kind of information is similar to what the EGAPP Working Group considers as part of its considerations of clinical utility and contextual issues.¹¹

CAN WE IMPLEMENT AN EVIDENCE-BASED TRIAGING OF GENOMIC TESTS THAT CAN DISTINGUISH SUBGROUPS WITHIN INSUFFICIENT EVIDENCE?

It is noteworthy that when evidence is insufficient, the use of informed or shared decision making is not new to the field of genetics. Indeed, for most rare, single-gene disorders, research on clinical utility has been limited or entirely lacking, but genetic counseling, using a model similar to that of informed/shared decision making has been effectively used for decades. For genetic tests for common diseases and those based on gene expressions and other complex biomarkers, we may be able to develop a similar transparent triaging approach for evidentiary classification of genomic applications. Similar to Veenstra et al.²

and the EGAPP Working Group, we think that not all “insufficient evidence” is created equal. Both Veenstra et al.² and the EGAPP working group use a combination of two dimensions to arrive at a recommendation, the first is the level of “certainty,” implying quantity and quality of the evidence, and the second is the magnitude of the “risk-benefit profile.” According to the methods paper of EGAPP,¹¹ the group is comfortable making a recommendation for or against the use of a genomic test only when the level of certainty is high or moderate. When the level of certainty is low, they return “insufficient evidence.” However, even then, they consider the importance of contextual factors such as the severity of the disorder, presence of therapeutic or diagnostic alternatives, current availability of the tests, costs, and other ethical and psychosocial issues to return a verdict of insufficient evidence that is “neutral,” “encouraging,” or “discouraging.” For example, although evidence was insufficient for the use of CYP450 testing for decision making in managing depression with selective serotonin reuptake inhibitors, the EGAPP Working Group discouraged use based on consideration of the contextual factors.¹³ A quantitative risk-benefit framework similar to what is proposed by Veenstra et al.² provides essential information in the recommendation development process than can facilitate the formal delineation of tests within the “insufficient evidence” category.

We propose that genomic tests be classified on the basis of available direct and indirect evidence into three tiers (implement in routine practice, do not use, and promote informed decision making) (Table 1). Beyond the ultimately desired binary outcome (Tiers 1 and 3), we must recognize that “insufficient evidence” (Tier 2) will be with us in public health and clinical genomics for decades to come. It is important, therefore, to make the most use of all the evidence available within the “insufficient evidence” category to guide practice and inform research. Consistent with approaches to decision making outlined by Veenstra et al.², EGAPP, and the USPSTF, it is possible to split the “insufficient evidence” category, what we call Tier 2, into two groups, 2a and 2b where 2b is similar to 3 in terms of the recommendation to not use or to discourage use and Tier 2a, for which the level of certainty is low but the preliminary risk-benefit profile analysis is favorable or promising. These tests should at least have established analytic and clinical validity even though final evidence of clinical utility may not be available. Although not definitive, the 2a category could merit a recommendation of “promoting informed decision-making” while conducting additional randomized clinical trials, comparative effectiveness studies, and/or public health surveillance on health outcomes.

Table 1 Evidence-based classification and examples of genomic tests for use in clinical practice

Tiers	1	2a	2b	3
Recommended action	Implement in practice	Informed decision making in practice	Do not use in practice	Do not use in practice
Evidence	Sufficient	Insufficient	Insufficient	Sufficient
Level of certainty	High/moderate	Low	Low	High/moderate
Risk benefit profile	Favorable	Favorable	Unfavorable or unknown	Unfavorable
Additional research needed	Implementation and outcomes research	CER, RCTs, IDM research	Basic, clinical, population research	Basic, clinical, population research
Examples	Lynch syndrome	Gene expression testing in tumors to improve outcomes in breast cancer	CYP450 testing before treating adult depression with SSRIs	Population screening for hereditary hemochromatosis
Potential health impact	3–5% of cases of colorectal cancer are due to Lynch	Help thousands of women diagnosed each year with early-stage breast cancer make decisions about treatment with chemotherapy	Enhance benefits and reduce side effects from SSRIs	Prevent illness and death from iron overload
Evidence Recommendation	EGAPP working group recommended that all new cases of colorectal cancer be screened; cascade testing of relatives can reduce deaths and illness ¹⁴	EGAPP working group found preliminary evidence of potential benefit to some who face decisions about treatment options but could not rule out potential harms to some women ¹⁶	EGAPP working group concluded there is insufficient evidence to support a recommendation for or against use; discouraged use of CYP450 testing ¹³	USPSTF concluded that harms outweighed benefits, because few with the genotype develop disease recommended against screening adult general population ²⁴
Action	Promote implementation through policy, education, and health systems interventions and developing clinical decision support tools	Promote informed decision making through policy, education, and health systems interventions and developing decision support tools	Discourage use in practice through policy, education and health systems interventions	Discourage use in practice through policy, education and health systems interventions

CER, comparative effectiveness research; RCTs, randomized controlled trials; IDM, research on how to conduct shared and informed decision making.

The use of informed or shared decision making is already well recognized in medicine and public health. There is a growing interest in patient education, patient–provider communication, and patient satisfaction with health care decision making. There is also a growing emphasis on informed choice by consumers; more patient involvement in health care decisions; greater quality and availability of evidence-based information on clinical options, including their pros and cons; increased understanding among both consumers and practitioners that many clinical decisions are not one-size-fits all and need to be less paternalistic and more sensitive to individual values.²⁰ With the proliferation of DTC genomic tests and other genome-based markers, for which there is clearly insufficient evidence of even clinical validity, we believe it is possible to use evidence-based approaches to put some of these tests into Tier 2b and, therefore, treat them similar to Tier 3 in terms of discouraging use in practice until further research is done.

To illustrate how the classification might work, we use in Table 1, four examples of genomic tests that have been evaluated by EGAPP and USPSTF and show the kind of actions that can be taken at the clinical and population levels to accompany evidence based classification. We use as Tier 1 example Lynch syndrome testing of all new colorectal cancer cases to reduce morbidity and mortality in first-degree relatives.¹⁴ For a Tier 2a example, we point to gene expression profiles to assess breast cancer recurrence risk and to target chemotherapy.¹⁶ We use testing for CYP450 genetic variants before treating adults with primary depression with selective serotonin reuptake inhibitors as a Tier 2b example.¹³ Finally, as a Tier 3 example, we use population screening for *HFE* gene mutations to prevent morbidity and mortality from iron overload.²⁴

These examples demonstrate the feasibility of a “binning process” to classify the evidence on genomic tests into categories but much more work will be needed to explore quantitative approaches to the classification of “insufficient evidence.”

CONCLUDING REMARKS

In summary, Veenstra et al.² propose the use of explicit and quantitative tools to the evaluation of genomic tests as they transition from research to practice. A recommendation matrix can be developed based on a quantitative assessment of certainty of the evidence and the assessment of the risk-benefit profile. We believe it is worthwhile to explore the development of a three-tier evidence-based classification for recommendation of genomic tests based on this approach (use in practice, promote informed decision making, and discourage use). We propose limiting informed decision making to tests for which there is sufficient information on analytic and clinical validity of the tests and for which the risk/benefit ratio is promising but not definitive. It may also be possible to refine this classification using types of tests and types of research needed (randomized controlled trials or other). All these approaches will depend strongly on a stakeholder-driven process to achieve buy-in, to refine such a schema, and to develop and apply methods for evaluation of emerging tests. We suggest that the recommendation of “promote informed decision-making” could emerge as a viable alternative to insufficient evidence in some cases. Although one can argue that informed decision making is ultimately needed for all tests, because of the numerous contextual issues involved in genomic testing, a too-liberal use of informed decision making also runs the risk of genomic medicine becoming an “evidence free” domain. Therefore, such a categorization should be used only when the evidence base can support it. Finally, we should strive to increase the amount and pace of randomized clinical trials, comparative effectiveness research,

and other novel modalities to assess the impact of using genomic tests on patients, families, and population outcomes to ensure the success of genomic medicine in the decades to come.

REFERENCES

- Petitti DB, Teutsch SM, Barton MB, et al. Update on the methods of the US Preventive Services Task Force: insufficient evidence. *Ann Intern Med* 2009;150:199–205.
- Veenstra DL, Roth JA, Garrison LP Jr, Ramsey SD, Burke W. A formal risk-benefit framework for genomic tests: facilitating the appropriate translation of genomics into clinical practice. *Genet Med*. In press.
- Evans JP, Burke W, Khoury M. The rules remain the same for genomic medicine: the case against “reverse genetic exceptionalism”. *Genet Med* 2010;12:342–343.
- Khoury MJ, Rich EC, Randhawa G, Teutsch SM, Niederhuber J. Comparative effectiveness research and genomic medicine: an evolving partnership for 21st century medicine. *Genet Med* 2009;11:707–711.
- Hunter DJ, Khoury MJ, Drazen JM. Letting the genome out of the bottle—will we get our wish? *N Engl J Med* 2008;358:105–107.
- Khoury MJ, Berg A, Coates RC, Evans J, Teutsch SM, Bradley LA. The evidence dilemma in genomic medicine. *Health Aff (Millwood)* 2008;27:1600–1611.
- Agency for Health Care Quality Research. The US Preventive Services Task Force (USPSTF), 2010. Available at: <http://www.ahrq.gov/clinic/uspstfix.htm>. Accessed August 8, 2010.
- National Institute for Health and Clinical Excellence (NICE), 2010. Available at: <http://www.nice.org.uk/>. Accessed August 8, 2010.
- The Cochrane Collaboration, 2010. Available at: <http://www.cochrane.org/>. Accessed August 8, 2010.
- The Guide to Community Preventive Services, 2010. Available at: <http://www.thecommunityguide.org/index.html>. Accessed August 8, 2010.
- Health Resources and Services Administration. Secretary’s Advisory Committee on Heritable Disorders in Newborns and Children (ACHDNC), 2010. Available at: <http://www.hrsa.gov/heritabledisorderscommittee/>. Accessed August 8, 2010.
- Evaluation of Genomic Applications in Practice and Prevention (EGAPP), 2010.
- Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group. Recommendations from the EGAPP Working Group: testing for cytochrome P450 polymorphisms in adults with nonpsychotic depression treated with selective serotonin reuptake inhibitors. *Genet Med* 2007;9:819–825.
- Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group. Recommendations from the EGAPP Working Group: genetic testing strategies in newly diagnosed individuals with colorectal cancer aimed at reducing morbidity and mortality from Lynch syndrome in relatives. *Genet Med* 2009;11:35–41.
- Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group. Recommendations from the EGAPP Working Group: can UGT1A1 genotyping reduce morbidity and mortality in patients with metastatic colorectal cancer treated with irinotecan? *Genet Med* 2009;11:15–20.
- Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group. Recommendations from the EGAPP Working Group: can tumor gene expression profiling improve outcomes in patients with breast cancer. *Genet Med* 2009;11:66–73.
- Wolf SH, Atkins D. The evolving role of prevention in healthcare: contributions of the U.S. Preventive Services Task Force. *Am J Prev Med* 2001;20:13–20.
- Teutsch SM, Bradley LA, Palomaki GE, et al. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) initiative: methods of the EGAPP Working Group. *Genet Med* 2009;11:3–14.
- Botkin JR, Teutsch SM, Kaye CI, et al. Outcomes of interest in evidence-based evaluations of genetic tests. *Genet Med* 2010;12:228–235.
- The Genomic Applications in Practice and Prevention Knowledge Base (GAPPKB), 2010. Available at: <http://www.hugenavigator.net/GAPPKB/home.do>. Accessed August 8, 2010.
- US Preventive Services Task Force. Screening for breast cancer: US Preventive Services Task Force recommendation statement. *Ann Intern Med* 2009;151:716–726, W-236.
- Sheridan SL, Harris RP, Woolf SH, Shared Decision-Making Workgroup of the U.S. Preventive Services Task Force. Shared decision-making about screening and chemoprevention. A suggested approach from the U.S. Preventive Services Task Force. *Am J Prev Med* 2004;26:56–66.
- Briss P, Rimer B, Reilley B, et al. and the Task Force on Community Preventive Services. Promoting informed decisions about cancer screening in communities and healthcare systems. *Am J Prev Med* 2004;26:67–80.
- US Preventive Services Task Force. Screening for hemochromatosis: recommendation statement. *Ann Intern Med* 2006;145:204–208.