

The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) initiative: methods of the EGAPP Working Group

Steven M. Teutsch, MD, MPH¹, Linda A. Bradley, PhD², Glenn E. Palomaki, BS³, James E. Haddow, MD³, Margaret Piper, PhD⁴, Ned Calonge, MD, MPH⁵, W. David Dotson, PhD^{2,6}, Michael P. Douglas, MS^{2,6}, and Alfred O. Berg, MD, MPH⁷, Chair, on behalf of the EGAPP Working Group

The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative, established by the National Office of Public Health Genomics at the Centers for Disease Control and Prevention, supports the development and implementation of a rigorous, evidence-based process for evaluating genetic tests and other genomic applications for clinical and public health practice in the United States. An independent, non-federal EGAPP Working Group (EWG), a multidisciplinary expert panel selects topics, oversees the systematic review of evidence, and makes recommendations based on that evidence. This article describes the EGAPP processes and details the specific methods and approaches used by the EWG. *Genet Med* 2009;11(1):3–14.

Key Words: Evidence-based medicine, review, systematic

The completion of the Human Genome Project has generated enthusiasm for translating genome discoveries into testing applications that have potential to improve health care and usher in a new era of “personalized medicine.”^{1–4} For the last decade however, questions have been raised about the appropriate evidentiary standards and regulatory oversight for this translation process.^{5–10} The US Preventive Services Task Force (USPSTF) was the first established national process to apply an evidence-based approach to the development of practice guidelines for genetic tests, focusing on *BRCA1/2* testing (to assess risk for heritable breast cancer) and on *HFE* testing for hereditary hemochromatosis.^{11,12} The Centers for Disease Control and Prevention-funded ACCE Project piloted an evidence evaluation framework of 44 questions that defines the scope of the review (i.e., disorder, genetic test, clinical scenario) and ad-

resses the previously proposed^{6,7} components of evaluation: Analytic and Clinical validity, Clinical utility and associated Ethical, legal and social implications. The ACCE Project examined available evidence on five genetic testing applications, providing evidence summaries that could be used by others to formulate recommendations.^{13–16} Systematic reviews on genetic tests have also been conducted by other groups.^{17–20}

Genetic tests tend to fit less well within “gold-standard” processes for systematic evidence review for several reasons.^{21–24} Many genetic disorders are uncommon or rare, making data collection difficult. Even greater challenges are presented by newly emerging genomic tests with potential for wider clinical use, such as genomic profiles that provide information on susceptibility for common complex disorders (e.g., diabetes, heart disease) or drug-related adverse events, and tests for disease prognosis.^{25,26} The actions or interventions that are warranted based on test results, and the outcomes of interest, are often not well defined. In addition, the underlying technologies are rapidly emerging, complex, and constantly evolving. Interpretation of test results is also complex, and may have implications for family members. Of most concern, the number and quality of studies are limited. Test applications are being proposed and marketed based on descriptive evidence and pathophysiologic reasoning, often lacking well-designed clinical trials or observational studies to establish validity and utility, but advocated by industry and patient interest groups.

THE EGAPP INITIATIVE

The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group (EWG) is an independent panel established in April, 2005, to develop a systematic process for evidence-based assessment that is specifically focused on genetic tests and other applications of genomic technology. Key objectives of the EWG are to develop a transparent, publicly accountable process, minimize conflicts of interest, optimize existing evidence review methods to address the challenges presented by complex and rapidly emerging genomic applications, and provide clear linkage between the scientific evidence and the subsequently developed EWG recommendation statements. The EWG is currently composed of 16 multidisciplinary experts in areas such as clinical practice, evidence-based medicine, genomics, public health, laboratory practice, epidemiology, economics, ethics, policy, and health technology assessment.²⁷ This nonfederal panel is supported by the EGAPP initiative launched in late 2004 by the National Office of Public Health Genomics (NOPHG) at the Centers for Disease Control and Prevention (CDC). In addition to supporting the activities of the EWG, EGAPP is developing data collection, synthesis, and review capacity to support timely and efficient translation of genomic applications into practice, evaluating the products and

From ¹Merck & Company, Inc., West Point, Pennsylvania; ²National Office of Public Health Genomics, CDC, Atlanta, Georgia; ³Department of Pathology and Laboratory Medicine (Research), The Warren Alpert Medical School of Brown University, Providence, Rhode Island; ⁴Blue Cross/Blue Shield Association Technology Evaluation Center, Chicago, Illinois; ⁵Colorado Department of Public Health and Environment, Denver, Colorado; ⁶McKing Consulting Corp., Atlanta, Georgia; and ⁷Department of Family Medicine, University of Washington, Seattle, Washington.

Linda A. Bradley, PhD, National Office of Public Health Genomics, Centers for Disease Control and Prevention, 4770 Buford Highway, MS K-89, Atlanta, GA 30341. E-mail: lbradley@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: Steven M. Teutsch is an employee, option and stockholder in Merck & Co., Inc. W. David Dotson is a former employee and current stockholder in Novo Nordisk, Inc.

Submitted for publication April 10, 2008.

Accepted for publication June 13, 2008.
DOI: 10.1097/GIM.0b013e318184137c

impact of the EWG’s pilot phase, and working with the EGAPP Stakeholders Group on topic prioritization, information dissemination, and product feedback.²⁸ The EWG is not a federal advisory committee, but rather aims to provide information to clinicians and other key stakeholders on the integration of genomics into clinical practice. The EGAPP initiative has no oversight or regulatory authority.

SCOPE AND SELECTION OF GENETIC TESTS AS TOPICS FOR EVIDENCE REVIEW

Much debate has centered on the definition of a “genetic test.” Because of the evolving nature of the tests and technologies, the EWG has adopted the broad view articulated in a recent report of the Secretary’s Advisory Committee on Genetics, Health, and Society¹⁰:

“A genetic test involves the analysis of chromosomes, deoxyribonucleic acid (DNA), ribonucleic acid (RNA), genes, or gene products (e.g., enzymes and other proteins) to detect heritable or somatic variations related to disease or health. Whether a laboratory method is considered a genetic test also depends on the intended use, claim or purpose of a test.”

Based on resource limitations, EGAPP focuses on tests having wider population application (e.g., higher disorder prevalence, higher frequency of test use), those with potential to impact clinical and public health practice (e.g., emerging prognostic and pharmacogenomic tests), and those for which there is significant demand for information. Tests currently eligible for EGAPP review include those used to guide intervention in symptomatic (e.g., diagnosis, prognosis, treatment) or asymptomatic individuals (e.g., disease screening), to identify individuals at risk for future disorders (e.g., risk assessment or susceptibility testing), or to predict treatment response or adverse events (e.g., pharmacogenomic tests) (Table 1). Though the

methods developed for systematic review are applicable, EGAPP is not currently considering diagnostic tests for rare single gene disorders, newborn screening tests, or prenatal screening and carrier tests for reproductive decision-making, as these tests are being addressed by other processes.^{10,29–39}

EGAPP-commissioned evidence reports and EWG recommendation statements are focused on patients seen in traditional primary or specialty care clinical settings, but may address other contexts, such as direct web-based offering of tests to consumers without clinician involvement (e.g., direct-to-consumer or DTC genetic testing). EWG recommendations may vary for different applications of the same test or for different clinical scenarios, and may address testing algorithms that include preliminary tests (e.g., family history or other laboratory tests that identify high risk populations).

Candidate topics (i.e., applications of genetic tests in specific clinical scenarios to be considered for evidence review) are identified through horizon scanning in the published and unpublished literature (e.g., databases, web postings), or nominated by EWG members, outside experts and consultants, federal agencies, health care providers and payers, or other stakeholders.⁴⁰ Like the USPSTF,²³ the EWG does not have an explicit process for ranking topics. EGAPP staff prepares background summaries on each potential topic which are reviewed and given preliminary priorities by an EWG Topics Subcommittee, based on specific criteria and aimed at achieving a diverse portfolio of topics that also challenge the evidence review methods (Table 2). Final selections are determined by vote of the full EWG. EGAPP is currently developing a more systematic and transparent process for prioritizing topics that is better informed by stakeholders.

REVIEW OF THE EVIDENCE

Evidence review strategies

When topics are selected for review by the EWG, CDC’s NOPHG commissions systematic reviews of the available evi-

Table 1 Categories of genetic test applications and some characteristics of how clinical validity and utility are assessed

Application of test	Clinical validity	Clinical utility
Diagnosis (symptomatic patient)	Association of marker with disorder	Improved clinical outcomes ^a —health outcomes based on diagnosis and subsequent intervention or treatment Availability of information useful for personal or clinical decision-making End of diagnostic odyssey
Disease screening (asymptomatic patient)	Association of marker with disorder	Improved health outcome based on early intervention for screen positive individuals to identify a disorder for which there is intervention or treatment, or provision of information useful for personal or clinical decision making
Risk assessment/susceptibility	Association of marker with future disorder (consider possible effect of penetrance)	Improved health outcomes based on prevention or early detection strategies
Prognosis of diagnosed disease	Association of marker with natural history benchmarks of the disorder	Improved health outcomes, or outcomes of value to patients, based on changes in patient management
Predicting treatment response or adverse events (pharmacogenomics)	Association of marker with a phenotype/metabolic state that relates to drug efficacy or adverse drug reactions	Improved health outcomes or adherence based on drug selection or dosage

^aClinical outcomes are the net health benefit (benefits and harms) for the patients and/or population in which the test is used.

Table 2 Criteria for preliminary ranking of topics

Criteria related to health burden	<p>What is the potential public health impact based on the prevalence/incidence of the disorder, the prevalence of gene variants, or the number of individuals likely to be tested?</p> <p>What is the severity of the disease?</p> <p>How strong is the reported relationship between a test result and a disease/drug response?</p> <p>Is there an effective intervention for those with a positive test or their family members?</p> <p>Who will use the information in clinical practice (e.g., healthcare providers, payers) and how relevant might this review be to their decision-making?</p>
Criteria related to practice issues	<p>What is the availability of the test in clinical practice?</p> <p>Is an inappropriate test use possible or likely?</p> <p>What is the potential impact of an evidence review or recommendations on clinical practice? On consumers?</p>
Other considerations	<p>How does the test add to the portfolio of EGAPP evidence based reviews? As a pilot project, EGAPP aims to develop a portfolio of evidence reviews that adequately tests the process and methodologies.</p> <p>Will it be possible to make a recommendation, given the body of data available? EGAPP is attempting to balance selection of somewhat established tests versus emerging tests for which insufficient evidence or unpublished data are more likely.</p> <p>Are there other practical considerations? For example, avoiding duplication of evidence reviews already underway by other groups.</p> <p>How does this test contribute to diversity in reviews? In what category is this test? As a pilot project, EGAPP aims to consider different categories of tests (e.g., pharmacogenomics or cancer), mutation types (e.g., inherited or somatic) or test types (e.g., predictive or diagnostic).</p>

dence. These reviews may include meta-analyses and economic evaluations. New topics are added on a phased schedule as funding and staff capacity allow. All EWG members, review team members, and consultants disclose potential conflicts of interest for each topic considered. Following the identification of the scope and the outcomes of interest for a systematic review, key questions and an analytic framework are developed by the EWG, and later refined by the review team in consultation with a technical expert panel (TEP). The EWG assigns members to serve on the TEP, along with other experts selected by those conducting the review; these members constitute the EWG “topic team” for that review. Based on the multidisciplinary nature of the panel, selection of EWG topic teams aims to include expertise in evidence-based medicine and scientific content.

For five of eight testing applications selected by the EWG to date, CDC-funded systematic evidence reviews have been conducted in partnership with the Agency for Healthcare Research and Quality (AHRQ) Evidence-based Practice Centers (EPCs).⁴¹ Based on expertise in conducting comprehensive, well-documented literature searches and evaluation, AHRQ EPCs represent an important resource for performing comprehensive reviews on applications of genomic technology. However, comprehensive reviews are time and resource intensive, and the numbers of relevant tests are rapidly increasing. Some tests have multiple applications and require review of more than one clinical scenario.^{7,10}

Consequently, the EWG is also investigating alternative strategies to produce shorter, less expensive, but no less rigorous, systematic reviews of the evidence needed to make decisions about immediate usefulness and highlight important gaps in knowledge. A key objective is to develop methods to support “targeted” or “rapid” reviews that are both timely and methodologically sound.^{13,17–20,42} Candidate topics for such reviews include situations when the published literature base is very limited, when it is possible to focus on a single evaluation

component (e.g., clinical validity) that is most critical for decision-making, and when information is urgently needed on a test with immediate potential for great benefit or harm. Three such targeted reviews are being coordinated by NOPHG-based EGAPP staff in collaboration with technical contractors, and with early participation of expert core consultants who can identify data sources and provide expert guidance on the interpretation of results.⁴³ Regardless of the source, a primary objective for all evidence reviews is that the final product is a comprehensive evaluation and interpretation of the available evidence, rather than summary descriptions of relevant studies.

Structuring the evidence review

“Evidence” is defined as peer-reviewed publications of original data or systematic review or meta-analysis of such studies; editorials and expert opinion pieces are not included.^{23,44} However, EWG methods allow for inclusion of peer-reviewed unpublished literature (e.g., information from Food and Drug Administration [FDA] Advisory Committee meetings), and for consideration on a case-by-case basis of other sources, such as review articles addressing relevant technical or contextual issues, or unpublished data. Topics are carefully defined based on the *medical disorder*, the *specific test (or tests)* to be used, and the specific *clinical scenario* in which it will be used.

The medical “disorder” (a term chosen as more encompassing than “disease”) should optimally be defined in terms of its clinical characteristics, rather than by the laboratory test being used to detect it. Terms such as condition or risk factor generally designate intermediate or surrogate outcomes or findings, which may be of interest in some cases; for example, identifying individuals at risk for atrial fibrillation as an intermediate outcome for preventing the clinical outcome of cardiogenic stroke. In pharmacogenomic testing, the disorder, or outcome of interest, may be a reduction in adverse drug events (e.g., avoiding severe neutropenia among cancer patients to be treated with

irinotecan via *UGT1A1* genotyping and dose reduction in those at high risk), optimizing treatment (e.g., adjusting initial warfarin dose using *CYP2C9* and *VKORC1* genotyping to more quickly achieve optimal anticoagulation in order to avoid adverse events), or more effectively targeting drug interventions to those patients most likely to benefit (e.g., herceptin for *HER2*-overexpressing breast cancers).

Characterizing the genetic test(s) is the second important step. For example, the American College of Medical Genetics defined the genetic testing panel for cystic fibrosis in the context of carrier testing as the 23 most common *CFTR* mutations (i.e., present at a population frequency of 0.1% or more) associated with classic, early onset cystic fibrosis in a US pan-ethnic study population. This allowed the subsequent review of analytic and clinical validity to focus on a relatively small subset of the 1000 or more known mutations.⁴⁵ Rarely, a nongenetic test may be evaluated, particularly if it is an existing alternative to mutation testing. An example would be biochemical testing for iron overload (e.g., serum transferrin saturation, serum ferritin) compared with *HFE* genotyping for identification of hereditary hemochromatosis.

A clear definition of the clinical scenario is of major importance, as the performance characteristics of a given test may vary depending on the intended use of the test, including the clinical setting (e.g., primary care, specialty settings), how the test will be applied (e.g., diagnosis or screening), and who will be tested (e.g., general population or selected high risk individuals). Preliminary tests should also be considered as part of the clinical scenario. For example, when testing for Lynch syndrome among newly diagnosed colorectal cancer cases, it may be too expensive to sequence two or more mismatch repair genes (e.g., *MLH1*, *MSH2*) in all patients. For this reason, preliminary tests, such as family history, microsatellite instability, or immunohistochemical testing, may be evaluated as strategies for selecting a smaller group of higher risk individuals to offer gene sequencing.

METHODS

Methods of the EWG for reviewing the evidence share many elements of existing processes, such as the USPSTF,²³ the AHRQ Evidence-based Practice Center Program,⁴⁶ the Centre for Evidence Based Medicine,⁴⁷ and others.^{44,48–53} These include the use of analytic frameworks with key questions to frame the evidence review; clear definitions of clinical and other outcomes of interest; explicit search strategies; use of hierarchies to characterize data sources and study designs; assessment of quality of individual studies and overall certainty of evidence; linkage of evidence to recommendations; and minimizing conflicts of interest throughout the process. Typically, however, the current evidence on genomic applications is limited to evaluating gene-disease associations, and is unlikely to include randomized controlled trials that evaluate test-based interventions and patient outcomes. Consequently, the EWG must rigorously assess the quality of observational studies, which may not be designed to address the questions posed.

In this new field, direct evidence to answer an overarching question about the effectiveness and value of testing is rarely available. Therefore, it is necessary to construct a chain of evidence, beginning with the technical performance of the test (analytic validity) and the strength of the association between a genotype and disorder of interest. The strength of this association determines the test's ability to diagnose a disorder, assess susceptibility or risk, or provide information on prognosis or variation in drug response (clinical validity). The final link is the

evidence that test results can change patient management decisions and improve net health outcomes (clinical utility).

To address some unique aspects of genetic test evaluation, the EWG has adopted several aspects of the ACCE model process, including formal assessment of analytic validity; use of unpublished literature for some evaluation components when published data are lacking or of low quality; consideration of ethical, legal, and social implications as integral to all components of evaluation; and use of questions from the ACCE analytic framework to organize collection of information.¹³ Important concepts that underlie the EGAPP process and add value include (1) providing a venue for multidisciplinary independent assessment of collected evidence; (2) conducting reviews that maintain a focus on medical outcomes that matter to patients, but also consider a range of specific family and societal outcomes when appropriate⁵⁴; (3) developing and optimizing methods for assessing individual study quality, adequacy of evidence for each component of the analytic framework, and certainty of the overall body of evidence; (4) focusing on summarization and synthesis of the evidence and identification of gaps in knowledge; and (5) ultimately, providing a foundation for evidentiary standards that can guide policy decisions. Although evidentiary standards will necessarily vary depending on test application (e.g., for diagnosis or to guide therapy) and the clinical situation, the methods and approaches described in this report are generally applicable; further refinement is anticipated as experience is gained.

The analytic framework and key questions

After the selection and structuring of the topic to be reviewed, the EWG Methods Subcommittee drafts an analytic framework for the defined topic that explicitly illustrates the clinical scenario, the intermediate and health outcomes of interest, and the key questions to be addressed. Table 1 provides generic examples of clinical scenarios. However, analytic frameworks for genetic tests differ based on clinical scenario, and must be customized for each topic. Figure 1 shows the example of an analytic framework used to develop the first EWG recommendation, *Testing for Cytochrome P450 Polymorphisms in Adults with Nonpsychotic Depression Prior to Treatment with Selective Serotonin Reuptake Inhibitors (SSRIs)*; numbers in the figure refer to the key questions listed in the legend.^{55,56}

The first key question is an over-arching question to determine whether there is direct evidence that using the test leads to clinically meaningful improvement in outcomes or is useful in medical or personal decision-making. In this case, EGAPP uses the USPSTF definition of direct evidence, "...a single body of evidence establishes the connection..." between the use of the genetic test (and possibly subsequent tests or interventions) and health outcomes.²³ Thus, the overarching question addresses clinical utility, and specific measures of the outcomes of interest. For genetic tests, such direct evidence on outcomes is most commonly not available or of low quality, so a "chain of evidence" is constructed using a series of key questions. EGAPP follows the convention that the chain of evidence is indirect if, rather than answering the overarching question, two or more bodies of evidence (linkages in the analytic framework) are used to connect the use of the test with health outcomes.^{23,57}

After the overarching question, the remaining key questions address the components of evaluation as links in a possible chain of evidence: *analytic validity* (technical test performance), *clinical validity* (the strength of association that determines the test's ability to accurately and reliably identify or predict the disorder of interest), and *clinical utility* (balance of benefits and harms when the test is used to

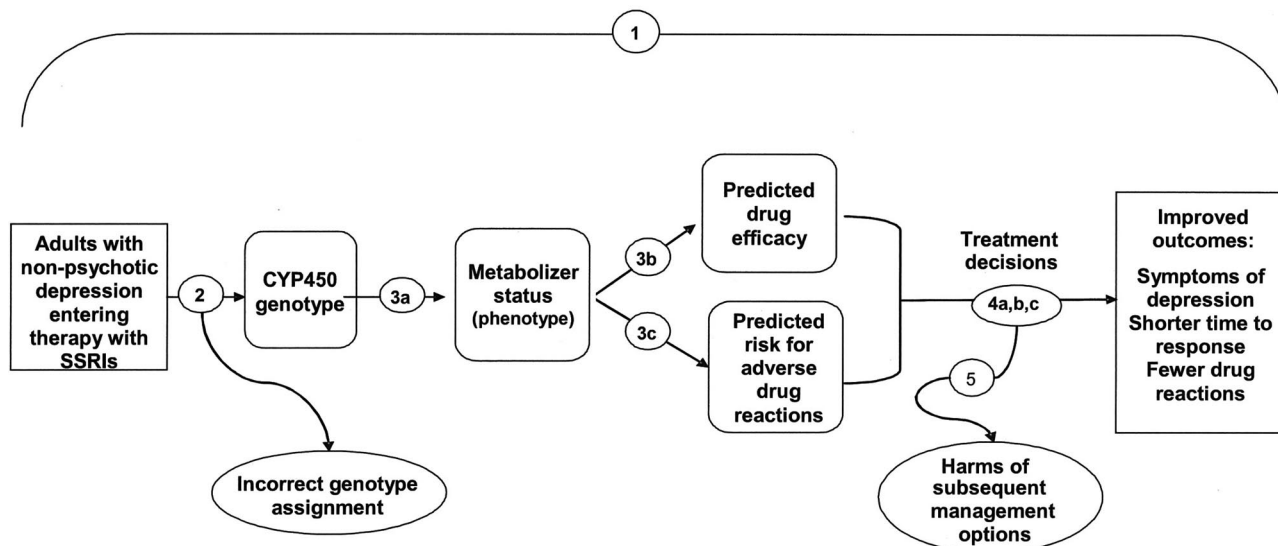


Fig. 1. Analytic framework and key questions for evaluating one application of a genetic test in a specific clinical scenario: *Testing for Cytochrome P450 Polymorphisms in Adults With Non-Psychotic Depression Treated With Selective Serotonin Reuptake Inhibitors (SSRIs)*; modified from reference 56. The numbers correspond to the following key questions:

1. Overarching question: Does testing for cytochrome P450 (*CYP450*) polymorphisms in adults entering selective serotonin reuptake inhibitor (SSRI) treatment for nonpsychotic depression lead to improvement in outcomes, or are testing results useful in medical, personal, or public health decision-making?
2. What is the *analytic validity* of tests that identify key *CYP450* polymorphisms?
3. *Clinical validity*: A, How well do particular *CYP450* genotypes predict metabolism of particular SSRIs? B, How well does *CYP450* testing predict drug efficacy? C, Do factors such as race/ethnicity, diet, or other medications, affect these associations?
4. *Clinical utility*: A, Does *CYP450* testing influence depression management decisions by patients and providers in ways that could improve or worsen outcomes? B, Does the identification of the *CYP450* genotypes in adults entering SSRI treatment for nonpsychotic depression lead to improved clinical outcomes compared to not testing? C, Are the testing results useful in medical, personal, or public health decision-making?
5. What are the harms associated with testing for *CYP450* polymorphisms and subsequent management options?

influence patient management). Determining whether a chain of indirect evidence can be applied to answer the overarching question requires consideration of the quality of individual studies, the adequacy of evidence for each link in the evidence chain, and the certainty of benefit based on the quantity (i.e., number and size) and quality (i.e., internal validity) of studies, the consistency and generalizability of results, and understanding of other factors or contextual issues that might influence the conclusions.^{23,57} The USPSTF has recently updated its methods and clarified its terminology.⁵⁷ Because this approach is both thoughtful and directly applicable to the work of EGAPP, the EWG has adopted the terminology; an additional benefit will be to provide consistency for shared audiences.

Evidence collection and assessment

The review team considers the analytic framework, key questions, and any specific methodological approaches proposed by the EWG. As previously noted, the report will focus on clinical factors (e.g., natural history of disease, therapeutic alternatives) and outcomes (e.g., morbidity, mortality, quality of life), but the EWG may request that other familial, ethical, societal, or intermediate outcomes also be considered for a specific topic.⁵⁴ The EWG may also request information

on other relevant factors (e.g., impact on management decisions by patients and providers) and contextual issues (e.g., cost-effectiveness, current use, or feasibility of use).

Methods for individual evidence reviews will differ in small ways based on the reviewers (AHRQ EPC or other review team), the strategy for review (e.g., comprehensive, targeted/rapid), and the topic. These differences will be transparent because all evidence reviews describe methods and follow the same general steps: framing the specific questions for review; gathering technical experts and reviewers; identifying data sources, searching for evidence using explicit strategies and study inclusion/exclusion criteria; specifying criteria for assessing quality of studies; abstracting data into evidence tables; synthesizing findings; and identifying gaps and making suggestions for future research.

All draft evidence reports are distributed to the TEP and other selected experts for technical review. After consideration of reviewer comments, EPCs provide a final report that is approved and released by AHRQ and posted on the AHRQ website; the EPC may subsequently publish a summary of the evidence. Non-EPC review teams submit final reports to CDC and the EWG, along with the comments from the technical reviewers and how they were addressed; the EWG approves the final report. Final evidence reports (or links to AHRQ

Table 3 Hierarchies of data sources and study designs for the components of evaluation

Level ^a	Analytic validity	Clinical validity	Clinical utility
1	Collaborative study using a large panel of well characterized samples Summary data from well-designed external proficiency testing schemes or interlaboratory comparison programs	Well-designed longitudinal cohort studies Validated clinical decision rule ^b	Meta-analysis of randomized controlled trials (RCT)
2	Other data from proficiency testing schemes Well-designed peer-reviewed studies (e.g., method comparisons, validation studies) Expert panel reviewed FDA summaries	Well-designed case-control studies	A single randomized controlled trial
3	Less well designed peer-reviewed studies	Lower quality case-control and cross-sectional studies Unvalidated clinical decision rule ^b	Controlled trial without randomization Cohort or case-control study
4	Unpublished and/or non-peer reviewed research, clinical laboratory, or manufacturer data Studies on performance of the same basic methodology, but used to test for a different target	Case series Unpublished and/or non-peer reviewed research, clinical laboratory or manufacturer data Consensus guidelines Expert opinion	Case series Unpublished and/or non-peer reviewed studies Clinical laboratory or manufacturer data Consensus guidelines Expert opinion

^aHighest level is 1.

^bA clinical decision rule is an algorithm leading to result categorization. It can also be defined as a clinical tool that quantifies the contributions made by different variables (e.g., test result, family history) in order to determine classification/interpretation of a test result (e.g., for diagnosis, prognosis, therapeutic response) in situations requiring complex decision-making.⁵⁵

reports) are posted on the www.egapreviews.org web site. When possible, a manuscript summarizing the evidence report is prepared to submit for publication along with the clinical practice recommendations developed by the EWG.⁵⁶

Grading quality of individual studies

Table 3 provides the hierarchies of data sources for analytic validity, and of study designs for clinical validity and utility, designated for all as Level 1 (highest) to Level 4. Table 4 provides a checklist of questions for assessing the quality of individual studies for each evaluation component based on the published literature.^{5,13,23,48,58,59} Different reviewers may provide a quality rating for individual studies that is based on specified criteria, or derived using a more quantitative algorithm. The EWG ranks individual studies as *Good*, *Fair*, or *Marginal* based on critical appraisal using the criteria in Tables 3 and 4. The designation *Marginal* (rather than *Poor*) acknowledges that some studies may not have been “poor” in overall design or conduct, but may not have been designed to address the specific key question in the evidence review.

Components of evaluation

Analytic validity

EGAPP defines the analytic validity of a genetic test as its ability to accurately and reliably measure the genotype (or analyte) of interest in the clinical laboratory, and in specimens representative of the population of interest.¹³ Analytic validity includes analytic sensitivity (detection rate), analytic specificity (1-false

positive rate), reliability (e.g., repeatability of test results), and assay robustness (e.g., resistance to small changes in preanalytic or analytic variables).¹³ As illustrated by the “ACCE wheel” figure (<http://www.cdc.gov/genomics/gtesting/ACCE.htm>), these elements of analytic validity are themselves integral elements in the assessment of clinical validity.^{13,42} Many evidence-based processes assume that evaluating clinical validity will address any analytic problems, and do not formally consider analytic validity.²³ The EWG has elected to pursue formal evaluation of analytic validity because genetic and genomic technologies are complex and rapidly evolving, and validation data are limited. New tests may not have been validated in multiple sites, for all populations of interest, or under routine clinical laboratory conditions over time. More importantly, review of analytic validity can also determine whether clinical validity can be improved by addressing test performance.

Tests kits or reagents that have been cleared or approved by the FDA may provide information on analytic validity that is publicly available for review (e.g., FDA submission summaries).⁶⁰ However, most currently available genetic tests are offered as laboratory developed tests not currently reviewed by the FDA, and information from other sources must be sought and evaluated. Different genetic tests may use a similar methodology, and information on the analytic validity of a common technology, as applied to genes not related to the review, may be informative. However, general information about the technology cannot be used as a substitute for specific information about the test under review. Based

Table 4 Criteria for assessing quality of individual studies (internal validity)⁵⁵

Analytic validity	Clinical validity	Clinical utility
<i>Adequate descriptions of the index test (test under evaluation)</i>	<i>Clear description of the disorder/phenotype and outcomes of interest</i>	<i>Clear description of the outcomes of interest</i>
Source and inclusion of positive and negative control materials	Status verified for all cases	What was the relative importance of outcomes measured; which were prespecified primary outcomes and which were secondary?
Reproducibility of test results	Appropriate verification of controls	<i>Clear presentation of the study design</i>
Quality control/assurance measures	Verification does not rely on <i>index test</i> result	Was there clear definition of the specific outcomes or decision options to be studied (clinical and other endpoints)?
<i>Adequate descriptions of the test under evaluation</i>	Prevalence estimates are provided	Was interpretation of outcomes/endpoints blinded?
Specific methods/platforms evaluated	<i>Adequate description of study design and test/methodology</i>	Were negative results verified?
Number of positive samples and negative controls tested	<i>Adequate description of the study population</i>	<i>Was data collection prospective or retrospective?</i>
<i>Adequate descriptions of the basis for the "right answer"</i>	Inclusion/exclusion criteria	If an experimental study design was used, were subjects randomized? Were intervention and evaluation of outcomes blinded?
Comparison to a "gold standard" referent test	Sample size, demographics	Did the study include comparison with current practice/empirical treatment (value added)?
Consensus (e.g., external proficiency testing)	Study population defined and representative of the clinical population to be tested	<i>Intervention</i>
Characterized control materials (e.g., NIST, sequenced)	Allele/genotype frequencies or analyte distributions known in general and subpopulations	What interventions were used?
<i>Avoidance of biases</i>	<i>Independent blind comparison with appropriate, credible reference standard(s)</i>	What were the criteria for the use of the interventions?
Blinded testing and interpretation	Independent of the test	<i>Analysis of data</i>
Specimens represent routinely analyzed clinical specimens in all aspects (e.g., collection, transport, processing)	Used regardless of test results	Is the information provided sufficient to rate the quality of the studies?
Reporting of test failures and uninterpretable or indeterminate results	Description of handling of indeterminate results and outliers	Are the data relevant to each outcome identified?
<i>Analysis of data</i>	Blinded testing and interpretation of results	Is the analysis or modeling explicit and understandable?
Point estimates of analytic sensitivity and specificity with 95% confidence intervals	<i>Analysis of data</i>	Are analytic methods prespecified, adequately described, and appropriate for the study design?
Sample size/power calculations addressed	Possible biases are identified and potential impact discussed	Were losses to follow-up and resulting potential for bias accounted for?
	Point estimates of clinical sensitivity and specificity with 95% confidence intervals	Is there assessment of other sources of bias and confounding?
	Estimates of positive and negative predictive values	Are there point estimates of impact with 95% CI?
		Is the analysis adequate for the proposed use?

NIST, National Institute of Standards and Quality.

on experience to date, access to specific expertise in clinical laboratory genetics and test development is important for effective review of analytic validity.

Table 3 (column 1) provides a quality ranking of data sources that are used to obtain unbiased and reliable information about analytic validity. The best information (quality Level 1) comes from collaborative studies using a single large, carefully selected panel of well-characterized samples (both cases and controls) that are blindly tested and reported, with the results independently analyzed. At this time, such studies are largely hypothetical, but an example that comes close is the Genetic Testing Quality Control Materials Program at CDC.⁶¹ As part of this program, samples precharacterized for specific genetic variants can be accessed from Coriell Cell Repositories (Camden, NJ) by other laboratories to perform in-house validation studies.⁶² Data from proficiency test-

ing schemes (Levels 1 or 2) can provide some information about all three phases of analytic validity (i.e., analytic, pre- and postanalytic), as well as interlaboratory and intermethod variability. ACCE questions 8 through 17 are helpful in ensuring that all aspects of analytic validity have been addressed.⁴²

Table 4 (column 1) lists additional criteria for assessing the quality of individual studies on analytic validity. Assessment of the overall quality of evidence for analytic validity includes consideration of the quality of studies, the quantity of data (e.g., number and size of studies, genes/alleles tested), and the consistency and generalizability of the evidence (also see Table 5, column 1). The consistency of findings can be assessed formally (e.g., by testing for homogeneity), or by less formal methods (e.g., providing a central estimate and range of values) when sufficient data are lacking. One or more internally valid studies do not necessarily

Table 5 Grading the quality of evidence for the individual components of the chain of evidence (key questions)⁵⁷

Adequacy of information to answer key questions	Analytic validity	Clinical validity	Clinical utility
Convincing	<i>Studies that provide confident estimates of analytic sensitivity and specificity using intended sample types from representative populations</i>	<i>Well-designed and conducted studies in representative population(s) that measure the strength of association between a genotype or biomarker and a specific and well-defined disease or phenotype</i>	<i>Well-designed and conducted studies in representative population(s) that assess specified health outcomes</i>
	Two or more Level 1 or 2 studies that are generalizable, have a sufficient number and distribution of challenges, and report consistent results	Systematic review/meta-analysis of Level 1 studies with homogeneity	Systematic review/meta-analysis of randomized controlled trials showing consistency in results At least one large randomized controlled trial (Level 2)
	One Level 1 or 2 study that is generalizable and has an appropriate number and distribution of challenges	Validated Clinical Decision Rule High quality Level 1 cohort study	
Adequate	Two or more Level 1 or 2 studies that	Systematic review of lower quality studies	Systematic review with heterogeneity
	Lack the appropriate number and/or distribution of challenges	Review of Level 1 or 2 studies with heterogeneity	One or more controlled trials without randomization (Level 3) Systematic review of Level 3 cohort studies with consistent results
	Are consistent, but not generalizable	Case/control study with good reference standards	
	Modeling showing that lower quality (Level 3, 4) studies may be acceptable for a specific well-defined clinical scenario	Unvalidated Clinical Decision Rule (Level 2)	
Inadequate	Combinations of higher quality studies that show important unexplained inconsistencies	Single case-control study Nonconsecutive cases	Systematic review of Level 3 quality studies or studies with heterogeneity
	One or more lower quality studies (Level 3 or 4)	Lacks consistently applied reference standards Single Level 2 or 3 cohort/case-control study	Single Level 3 cohort or case-control study
	Expert opinion	Reference standard defined by the test or not used systematically Study not blinded	Level 4 data
		Level 4 data	

provide sufficient information to conclude that analytic validity has been established for the test. Supporting the use of a test in routine clinical practice requires data on analytic validity that are generalizable to use in diverse “real world” settings.

Clinical validity

EGAPP defines the clinical validity of a genetic test as its ability to accurately and reliably predict the clinically defined disorder or phenotype of interest. Clinical validity encompasses clinical sensitivity and specificity (integrating analytic validity), and predictive values of positive and negative tests that take into account the disorder prevalence (the proportion of individuals in the selected setting who have, or will develop, the phenotype/clinical disorder of interest). Clinical validity may also be affected by reduced penetrance (i.e., the proportion of individuals with a disease-related genotype or mutation who develop disease), variable expressivity (i.e., variable severity of disease among individuals with the same genotype), and other genetic (e.g., variability in allele/genotype frequencies or gene-disease association in racial/ethnic subpopula-

tions) or environmental factors. ACCE questions 18 through 25 are helpful in organizing information on clinical validity.⁴²

Table 3 (column 2) provides a hierarchy of study designs for assessing quality of individual studies.^{13,23,44,46–48,50,53,63} Published checklists for reporting studies on clinical validity are reasonably consistent, and Table 4 (column 2) provides additional criteria adopted for grading the quality of studies (e.g., execution, minimizing bias).^{5,13,23,44,46–51,53,58,59,63} As with analytic validity, the important characteristics defining overall quality of evidence on clinical validity include the number and quality of studies, the representativeness of the study population(s) compared with the population(s) to be tested, and the consistency and generalizability of the findings (Table 5). The quantity of data includes the number of studies, and the number of total subjects in the studies. The overall consistency of clinical validity estimates can be determined by formal methods such as meta-analysis. Minimally, estimates of clinical sensitivity and specificity should include confidence intervals.⁶³ In pilot studies, initial estimates of clinical validity

may be derived from small data sets focused on individuals known to have, versus not have, a disorder, or from case/control studies that may not represent the wide range or frequency of results that will be found in the general population. Although important to establish proof of concept, such studies are insufficient evidence for clinical application; additional data are needed from the entire range of the intended clinical population to reliably quantify clinical validity before introduction.

Clinical utility

EGAPP defines the clinical utility of a genetic test as the evidence of improved measurable clinical outcomes, and its usefulness and added value to patient management decision-making compared with current management without genetic testing. If a test has utility, it means that the results (positive or negative) provide information that is of value to the person, or sometimes to the individual’s family or community, in making decisions about effective treatment or preventive strategies. Clinical utility encompasses effectiveness (evidence of utility in real clinical settings), and the net benefit (the balance of benefits and harms). Frequently, it also involves assessment of efficacy (evidence of utility in controlled settings like a clinical trial).

Tables 3 and 4 (column 3) provide the hierarchy of study designs for clinical utility, and other criteria for grading the internal validity of studies (e.g., execution, minimizing bias) adopted from other published approaches.^{13,23,46–48,57} Paralleling the assessment of analytic and clinical validity, the three important quality characteristics for clinical utility are quality of individual studies and the overall body of evidence, the quantity of relevant data, and the consistency and generalizability of the findings (Table 5). Another criterion to be considered is whether implementation of testing in different settings, such as clinician ordered versus direct-to-consumer, could lead to variability in health outcomes.

Grading the quality of evidence for the individual components in the chain of evidence (key questions)

Table 5 provides criteria for assessing the quality of the body of evidence for the individual components of evaluation, ana-

lytic validity (column 2), clinical validity (column 3), and clinical utility (column 4).^{23,44,47,48,64} The adequacy of the information to answer the key questions related to each evaluation component is classified as *Convincing*, *Adequate*, or *Inadequate*. This information is critical to assess the “strength of linkages” in the chain of evidence.⁵⁷ The intent of this approach is to minimize the risk of being wrong in the conclusions derived from the evidence. When the quality of evidence is *Convincing*, the observed estimate or effect is likely to be real, rather than explained by flawed study methodology; when *Adequate*, the observed results may be influenced by such flaws. When the quality of evidence is *Inadequate*, the observed results are more likely to be the result of flaws in study methodology rather than an accurate assessment; availability of only *Marginal* quality studies always results in *Inadequate* quality.

Based on the evidence available, the overall level of certainty of net health benefit is categorized as *High*, *Moderate*, or *Low*.⁵⁷ *High* certainty is associated with consistent and generalizable results from well-designed and conducted studies, making it unlikely that estimates and conclusions will change based on future studies. When the level of certainty is *Moderate*, some data are available, but limitations in data quantity, quality, consistency, or generalizability reduce confidence in the results, and, as more information becomes available, the estimate or effect may change enough to alter the conclusion. *Low* certainty is associated with insufficient or poor quality data, results that are not consistent or generalizable, or lack of information on important outcomes of interest; as a result, conclusions are likely to change based on future studies.

Translating evidence into recommendations

Based on the evidence report, the EWG’s assessment of the magnitude of net benefit and the certainty of evidence, and consideration of other clinical and contextual issues, the EWG formulates clinical practice recommendations (Table 6). Although the information will have value to other stakeholders, the primary intended audience for the content and format of the recommendation statement is clinicians. The information is intended to provide transparent, authoritative advice, inform targeted research agendas, and underscore the increasing need

Table 6 Recommendations based on certainty of evidence, magnitude of net benefit, and contextual issues

Level of Certainty	Recommendation
High or moderate	Recommend for if the magnitude of net benefit is <i>Substantial</i> , <i>Moderate</i> , or <i>Small</i> ^a , unless additional considerations warrant caution. Consider the importance of each relevant contextual factor and its magnitude or finding.
	Recommend against if the magnitude of net benefit is <i>Zero</i> or there are net harms. Consider the importance of each relevant contextual factor and its magnitude or finding.
Low	Insufficient evidence if the evidence for clinical utility or clinical validity is insufficient in quantity or quality to support conclusions or make a recommendation. Consider the importance of each contextual factor and its magnitude or finding. Determine whether the recommendation should be <i>Insufficient</i> (neutral), <i>Insufficient</i> (encouraging), or <i>Insufficient</i> (discouraging). Provide information on key information gaps to drive a research agenda.

^aCategories for the “magnitude of effect” or “magnitude of net benefit” used are *Substantial*, *Moderate*, *Small*, and *Zero*.⁵⁷

for translational research that supports the appropriate transition of genomic discoveries to tests, and then to specific clinical applications that will improve health or add other value in clinical practice.

Key factors considered in the development of a recommendation are the relative importance of the outcomes selected for review, the benefits (e.g., improved clinical outcome, reduction of risk) that result from the use of the test and subsequent actions or interventions (or if not available, maximum potential benefits), the harms (e.g., adverse clinical outcome, increase in risk or burden) that result from the use of the test and subsequent actions/interventions (or if not available, largest potential harms), and the efficacy and effectiveness of the test and follow-up compared with currently used interventions (or doing nothing). Simple decision models or outcomes tables may be used to assess the magnitudes of benefits and harms, and estimate the net effect. Consistent with the terminology used by the USPSTF, the magnitude of net benefit (benefit minus harm) may be classified as *Substantial*, *Moderate*, *Small*, or *Zero*.⁵⁷

Considering contextual factors

Contextual issues include clinical factors (e.g., severity of disorder, therapeutic alternatives), availability of diagnostic alternatives, current availability and use of the test, economics (e.g., cost, cost-effectiveness, and opportunity costs), and other ethical and psychosocial considerations (e.g., insurability, family factors, acceptability, equity/fairness). Cost-effectiveness analysis is especially important when a recommendation for testing is made. Contextual issues that are not included in preparing EGAPP recommendation statements are values or preferences, budget constraints, and precedent. Societal perspectives on whether use of the test in the proposed clinical scenario is ethical are explored before commissioning an evidence review.

The ACCE analytic framework considers as part of clinical utility the assessment of a number of additional elements related to the integration of testing into routine practice (e.g., adequate facilities/resources to support testing and appropriate follow-up, plan for monitoring the test in practice, availability of validated educational materials for providers and consumers).¹³ The EWG considers that most of these elements constitute information that should not be included in the consideration of clinical utility, but may be considered as contextual factors in developing recommendation statements and in translating recommendations into clinical practice.

Recommendation language

Standard EGAPP language for recommendation statements uses the terms: *Recommend For*, *Recommend Against*, or *Insufficient Evidence* (Table 6). Because the types of emerging genomic tests addressed by EGAPP are more likely to have findings of *Insufficient Evidence*, three additional qualifiers may be added. Based on the existing evidence and consideration of contextual issues and modeling, *Insufficient Evidence* could be considered “Neutral” (not possible to predict with current evidence), “Discouraging” (discouraged until specific gaps in knowledge are filled or not likely to meet evidentiary standards even with further study), and “Encouraging” (likely to meet evidentiary standards with further studies or reasonable to use in limited situations based on existing evidence while additional evidence is gathered).

As a hypothetical example of how the various components of the review are brought together to reach a conclusion, consider the model of a pharmacogenetic test proposed for screening individuals who are entering treatment with a specific drug. The

intended use is to identify individuals who are at risk for a serious adverse reaction to the drug. The analytic validity and clinical validity of the test are established and adequately high. However, the specific adverse outcomes of interest are often clinically diagnosed and treated as part of routine management, and clinical studies have not been conducted to show the incremental benefit of the test in improving patient outcomes. Because there is no evidence to support improvement in health outcome or other benefit of using the test (e.g., more effective, more acceptable to patients, or less costly), the EWG would consider the recommendation to be *Insufficient Evidence* (neutral). In a second scenario, a genetic test is proposed for testing patients with a specific disorder to provide information on prognosis and treatment. Clinical trials have provided good evidence for benefit to a subset of patients based on the test results, but more studies are needed to determine the validity and utility of testing more generally. The EWG is likely to consider the recommendation to be *Insufficient Evidence* (encouraging).

Products and review

Draft evidence reports are distributed by the EPC or other contractor for expert peer-review. Objectives for peer review of draft evidence reports are to ensure accuracy, completeness, clarity, and organization of the document; assess modeling, if present, for parameters, assumptions and clinical relevance; and to identify scientific or contextual issues that need to be addressed or clarified in the final evidence report. In general, the selection of reviewers is based on expertise, with consideration given to potential conflicts of interest.

When a final evidence report is received by the EWG, a writing team begins development of the recommendation statement. Technical comments are solicited from test developers on the evidence report’s accuracy and completeness, and are considered by the writing team. The recommendation statement is intended to summarize current knowledge on the validity and utility of an intended use of a genetic test (what we know and do not know), consider contextual issues related to implementation, provide guidance on appropriate use, list key gaps in knowledge, and suggest a research agenda. Following acceptance by the full EWG, the draft EGAPP recommendation statement is distributed for comment to peer reviewers selected from organizations expected to be impacted by the recommendation, the EGAPP Stakeholders Group, and other key target audiences (e.g., health care payers, consumer organizations). The objectives of this peer review process are to ensure the accuracy and completeness of the evidence summarized in the recommendation statement and the transparency of the linkage to the evidence report, improve the clarity and organization of information, solicit feedback from different perspectives, identify contextual issues that have not been addressed, and avoid unintended consequences. Final drafts of recommendation statements are approved by the EWG and submitted for publication in *Genetics in Medicine*. Once published, the journal provides open access to these documents, and the link is also posted on the www.egappreviews.org web site. Announcements of recommendation statements are distributed by email to a large number of stakeholders and the media. The newly established EGAPP Stakeholders Group will advise on and facilitate dissemination of evidence reports and recommendation statements.

Summary

This document describes methods developed by the EWG for establishing a systematic, evidence-based assessment process that is specifically focused on genetic tests and other applica-

tions of genomic technology. The methods aim for transparency, public accountability, and minimization of conflicts of interest, and provide a framework to guide all aspects of genetic test assessment, beginning with topic selection and concluding with recommendations and dissemination. Key objectives are to optimize existing evidence review methods to address the challenges presented by complex and rapidly emerging genomic applications, and to establish a clear linkage between the scientific evidence, the conclusions/recommendations, and the information that is subsequently disseminated.

In combining elements from other internationally recognized assessment schemes in its methods, the EWG seeks to maintain continuity in approach and nomenclature, avoid confusion in communication, and capture existing expertise and experience. The panel's methods differ from others in some respects, however, by calling for formal assessment of analytic validity (in addition to clinical validity and clinical utility) in its evidence reviews, and including (on a selective basis) nontraditional sources of information such as gray literature, unpublished data, and review articles that address relevant technical or contextual issues. The methods and process of the EWG remain a work in progress and will continue to evolve as knowledge is gained from each evidence review and recommendation statement.

Future challenges include modifying current methods to achieve more rapid, less expensive, and targeted evidence reviews for test applications with limited literature, without sacrificing the quality of the answers needed to inform practice decisions and research agendas. A more systematic horizon scanning process is being developed to identify high priority topics more effectively, in partnership with the EGAPP Stakeholders Group and other stakeholders. Additional partnerships will need to be created to develop evidentiary standards and build additional evidence review capacity, nationally. Finally, the identification of specific gaps in knowledge in the evidence offers the opportunity to raise awareness among researchers, funding entities, and review panels, and thereby focus future translation research agendas.

ACKNOWLEDGMENTS

Members of the EGAPP Working Group are Alfred O. Berg, MD, MPH, Chair; Katrina Armstrong, MD, MSCE; Jeffrey Botkin, MD, MPH; Ned Calonge, MD, MPH; James E. Haddow, MD; Maxine Hayes, MD, MPH; Celia Kaye, MD, PhD; Kathryn A. Phillips, PhD; Margaret Piper, PhD, MPH; Sue Richards, PhD; Joan A. Scott, MS, CGC; Ora Strickland, PhD; Steven Teutsch, MD, MPH.

REFERENCES

- Burke W, Psaty BM. Personalized medicine in the era of genomics. *JAMA* 2007;298:1682–1684.
- Gupta P, Lee KH. Genomics and proteomics in process development: opportunities and challenges. *Trends Biotechnol* 2007;25:324–330.
- Topol EJ, Murray SS, Frazer KA. The genomics gold rush. *JAMA* 2007;298:218–221.
- Feero WG, Guttmacher AE, Collins FS. The genome gets personal—almost. *JAMA* 2008;299:1351–1352.
- Burke W, Atkins D, Gwinn M, et al. Genetic test evaluation: information needs of clinicians, policy makers, and the public. *Am J Epidemiol* 2002;156:311–318.
- Holtzman NA, Watson MS. Promoting safe and effective genetic testing in the United States. Final Report of the National Institute of Health - Department of Energy (DOE) Task Force on Genetic Testing. Available at: <http://www.genome.gov/10001733>. Accessed December 12, 2007.
- Secretary's Advisory Committee on Genetics, Health and Society. Enhancing oversight of genetic tests: recommendations of the SACGT. Available at: http://www4.od.nih.gov/oba/sacgt/reports/oversight_report.pdf. Accessed December 12, 2007.
- Huang A. Genetics & Public Policy Center, Issue Briefs. Who regulates genetic tests? Available at: http://www.dnapolicy.org/policy.issue.php?action=detail&issuebrief_id=10. Accessed December 12, 2007.
- Secretary's Advisory Committee on Genetics, Health and Society. Coverage and reimbursement of genetic tests and services. Available at: http://www4.od.nih.gov/oba/SACGHS/reports/CR_report.pdf. Accessed December 12, 2007.
- U.S. System of Oversight of Genetic Testing. A response to the charge of the secretary of health and human services. Report of the Secretary's Advisory Committee on Genetics, Health, and Society. April, 2008. Department of Health and Human Services. Available at: http://www4.od.nih.gov/oba/sacghs/reports/SACGHS_oversight_report.pdf.
- USPSTF. Genetic risk assessment and BRCA mutation testing for breast and ovarian cancer susceptibility: recommendation statement. *Ann Intern Med* 2005;143:355–361.
- Whitlock EP, Garlitz BA, Harris EL, Beil TL, Smith PR. Screening for hereditary hemochromatosis: a systematic review for the U.S. Preventive Services Task Force. *Ann Intern Med* 2006;145:209–223.
- Haddow J, Palomaki G. ACCE: a model process for evaluating data on emerging genetic tests. In: Khoury M, Little J, Burke W, editors. Human genome epidemiology: a scientific foundation for using genetic information to improve health and prevent disease. New York: Oxford University Press, 2003:217–233.
- Palomaki GE, Haddow J, Bradley L, Fitzsimmons SC. Updated assessment of cystic fibrosis mutation frequencies in non-Hispanic Caucasians. *Genet Med* 2002;4: 90–94.
- Palomaki G, Bradley L, Richards C, Haddow J. Analytic validity of cystic fibrosis testing: a preliminary estimate. *Genet Med* 2003;5:15–20.
- Palomaki G, Haddow J, Bradley L, Richards CS, Stenzel TT, Grody WW. Estimated analytic validity of HFE C282Y mutation testing in population screening: the potential value of confirmatory testing. *Genet Med* 2003;5: 440–443.
- Gudgeon J, McClain M, Palomaki G, Williams M. Rapid ACCE: experience with a rapid and structured approach for evaluating gene-based testing. *Genet Med* 2007;9:473–478.
- McClain M, Palomaki G, Piper M, Haddow J. A rapid ACCE review of *CYP2C9* and *VKORC1* allele testing to inform warfarin dosing in adults at elevated risk for thrombotic events to avoid serious bleeding. *Genet Med* 2008;10:89–98.
- Piper MA. Blue Cross and Blue Shield Special Report. Genotyping for cytochrome P450 polymorphisms to determine drug-metabolizer status. Available at: www.bcbs.com/tec/Vol19/19_09.pdf. Accessed July 6, 2006.
- Piper MA. Blue Cross Blue Shield TEC Assessment. Gene expression profiling of breast cancer to select women for adjuvant chemotherapy. Available at: http://www.bcbs.com/blueresources/tec/vols/22/22_13.pdf. Accessed August 18, 2008.
- Briss P, Zaza S, Pappaioanou M, et al. Developing an evidence-based Guide to Community Preventive Services—methods. The Task Force on Community Preventive Services. *Am J Prev Med* 2000;18:35–43.
- Briss P, Brownson R, Fielding J, Zaza S. Developing and using the Guide to Community Preventive Services: lessons learned about evidence-based public health. *Annu Rev Public Health* 2008;25:281–302.
- Harris R, Helfand M, Woolf S, et al. Current methods of the US Preventive Services Task Force: a review of the process. *Am J Prev Med* 2001;20:21–35.
- Zaza S, Wright-De A, Briss P, et al. Data collection instrument and procedure for systematic reviews in the Guide to Community Preventive Services. Task Force on Community Preventive Services. *Am J Prev Med* 2000;18: 44–74.
- Hunter D, Khoury M, Drazen J. Letting the genome out of the bottle—will we get our wish? *N Engl J Med* 2008;358:105–107.
- Kamerow D. Waiting for the genetic revolution. *BMJ* 2008;336:22.
- EGAPPreviews.org Homepage. Available at: <http://www.egappreviews.org/>. Accessed March 20, 2008.
- National Office of Public Health Genomics, CDC Web site. Available at: <http://www.cdc.gov/genomics>. Accessed March 20, 2008.
- Maddalena A, Bale S, Das S, et al. Technical standards and guidelines: molecular genetic testing for ultra-rare disorders. *Genet Med* 2005;7:571–583.
- Faucett WA, Hart S, Pagon RA, Neall LF, Spinella G. A model program to increase translation of rare disease genetic tests: collaboration, education, and test translation program. *Genet Med* 2008;10:343–348.
- Gross SJ, Pletcher BA, Monaghan KG. Carrier screening in individuals of Ashkenazi Jewish descent. *Genet Med* 2008;10:54–56.
- Driscoll DA, Gross SJ. First trimester diagnosis and screening for fetal aneuploidy. *Genet Med* 2008;10:73–75.
- Grody WW, Cutting GR, Klinger KW, et al. Laboratory standards and guidelines for population-based cystic fibrosis carrier screening. *Genet Med* 2001;3:149–154.
- Richards CS, Bradley LA, Amos J, et al. Standards and guidelines for CFTR mutation testing. *Genet Med* 2002;4:379–391.
- Evidence-based evaluation and decision process for the Advisory Committee on Heritable Disorders and Genetic Diseases in Newborn and Children: a work-

- group meeting summary. October 23, 2006. Advisory Committee on Heritable Disorders and Genetic Diseases in Newborns and Children. Available at: <ftp://ftp.hrsa.gov/mchb/genetics/reports/MeetingSummary23Oct2006.pdf>. Accessed June 10, 2008.
36. Advisory Committee on Heritable Disorders and Genetic Diseases in Newborns and Children. U.S. Department of Health and Human Services, Health Resources and Services Administration. Available at: <http://www.hrsa.gov/heritabledisorderscommittee/>. Accessed June 10, 2008.
 37. National Newborn Screening and Genetics Resource Center. Available at: <http://genes-r-us.uthscsa.edu/>. Accessed June 10, 2008.
 38. American College of Obstetricians and Gynecologists News Release. ACOG's screening guidelines on chromosomal abnormalities: what they mean to patients and physicians. Available at: http://www.acog.org/from_home/publications/press_releases/nr05-07-07-1.cfm. Accessed June 11, 2008.
 39. Genetic advances and the rarest of rare diseases: genetics in medicine focuses on genetics and rare diseases. Available at: <http://www.acmg.net/AM/Template.cfm?Section=Home&Template=/CM/HTMLDisplay.cfm&ContentID=3049>. Accessed June 11, 2008.
 40. National Office of Public Health Genomics—EGAPP Stakeholders Group. Available at: http://www.cdc.gov/genomics/gtesting/egapp_esg.htm. Accessed January 15, 2008.
 41. Agency for Healthcare Research and Quality Evidence-based Practice Centers. Available at: <http://www.ahrq.gov/clinic/epc/>. Accessed January 15, 2008.
 42. National Office of Public Health Genomics, CDC. ACCE model system for collecting, analyzing and disseminating information on genetic tests. Available at: <http://www.cdc.gov/genomics/gtesting/ACCE.htm>. Accessed December 12, 2007.
 43. EGAPP Reviews.org Topics. Available at: <http://www.egappreviews.org/workingrp/topics.htm>. Accessed October 9, 2007.
 44. Ebell M, Siwek J, Weiss B, et al. Strength of recommendation taxonomy (SORT): a patient-centered approach to grading evidence in the medical literature. *J Am Board Fam Pract* 2004;17:59–67.
 45. ACCE. Population-based prenatal screening for cystic fibrosis via carrier testing—introduction. Available at: <http://www.cdc.gov/genomics/gtesting/ACCE/FBR/CF/CFIntro.htm>. Accessed January 16, 2008.
 46. AHRQ. Systems to rate the strength of scientific evidence. Evidence Report/Technology Assessment. Available at: <http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=hstat1.chapter.70996>. Accessed December 12, 2007.
 47. Centre for Evidence Based Medicine. Levels of evidence and grades of recommendation. Available at: http://www.cebm.net/levels_of_evidence.asp. Accessed December 12, 2007.
 48. Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. *BMJ* 2004;328:1490.
 49. Bossuyt P, Reitsma J, Bruns D, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Ann Intern Med* 2003;138: 40–44.
 50. Deeks JJ. Systematic reviews in health care: systematic reviews of evaluations of diagnostic and screening tests. *BMJ* 2001;323:157–162.
 51. Tatsioni A, Zarin D, Aronson N, et al. Challenges in systematic reviews of diagnostic technologies. *Ann Intern Med* 2005;142:1048–1055.
 52. US Food and Drug Administration. Draft guidance for industry, clinical laboratories, and FDA staff. In vitro diagnostic multivariate index assays. Available at: <http://www.fda.gov/cdrh/oivd/guidance/1610.pdf>. Accessed January 15, 2008.
 53. Whiting P, Rutjes A, Reitsma J, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003; 3:25.
 54. Botkin JR, Teutsch SM, Kaye CI, et al. Examples of types of health-related outcomes. Table 5–2 in: U.S. System of Oversight of Genetic Testing: a response to the Charge of the Secretary of Health and Human Services. Report of the Secretary's Advisory Committee on Genetics, Health, and Society, 2008:122–123. Available at: http://www4.od.nih.gov/oba/SACGHS/reports/SACGHS_oversight_report.pdf. Accessed June 12, 2008.
 55. Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group. Recommendations from the EGAPP Working Group: testing for cytochrome P450 (CYP450) polymorphisms in adults with nonpsychotic depression treated with selective serotonin reuptake inhibitors. *Genet Med* 2007;9:819–825.
 56. Matchar DB, Thakur ME, Grossman I, et al. Testing for cytochrome P450 polymorphisms in adults with non-psychotic depression treated with selective serotonin reuptake inhibitors (SSRIs). Evidence Report/Technology Assessment No. 146. (Prepared by the Duke Evidence-based Practice Center under Contract No. 290-02-0025.) AHRQ Publication No. 07-E002. Rockville, MD: Agency for Healthcare Research and Quality. Available at: <http://www.ahrq.gov/downloads/pub/evidence/pdf/cyp450/cyp450.pdf>. Accessed January 23, 2007.
 57. Sawaya GF, Guirguis-Blake J, LeFevre M, et al. Update on methods of the U.S. Preventive Services Task Force: estimating certainty and magnitude of net benefit. *Ann Intern Med* 2007;147:871–875.
 58. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;282:1061–1066.
 59. Little J, Bradley L, Bray M, et al. Reporting, appraising, and integrating data on genotype prevalence and gene-disease associations. *Am J Epidemiol* 2002;156:300–310.
 60. U.S. Food and Drug Administration. 510(k) Substantial equivalence determination decision summary for Roche AmpliChip CYP450 microarray for identifying CYP2D6 genotype (510(k) Number k042259). Available at: <http://www.fda.gov/cdrh/reviews/k042259.pdf>. Accessed April 19, 2006.
 61. Genetic Testing Reference Materials Coordination Program (GeT-RM)—Home. Available at: <http://www.cdc.gov/dls/genetics/rmmaterials/default.aspx>. Accessed January 17, 2008.
 62. Cell lines typed for CYP2C9 and VKORC1 alleles. Available at: http://www.cdc.gov/dls/genetics/rmmaterials/pdf/CYP2C9_VKORC1.pdf. Accessed January 17, 2008.
 63. American College of Medical Genetics. C8—test validation. Standards and guidelines for clinical genetics laboratories: general policies. Available at: http://www.acmg.net/Pages/ACMG_Activities/stds-2002/c.htm. Accessed December 12, 2007.
 64. Clinical evidence: the international source of the best available evidence for effective health care. Available at: <http://clinicalevidence.bmj.com/ceweb/about/index.jsp>. Accessed January 16, 2008.