

# Chapter 11

## Challenges in and Principles for Conducting Systematic Reviews of Genetic Tests Used as Predictive Indicators

Daniel E. Jonas, M.D., M.P.H., University of North Carolina

Timothy J. Wilt, M.D., M.P.H., Department of Veterans Affairs Health Care System,  
Minneapolis; University of Minnesota

Brent C. Taylor, Ph.D., M.P.H., Department of Veterans Affairs Health Care System,  
Minneapolis; University of Minnesota

Tania M. Wilkins, M.S., University of North Carolina

David B. Matchar, M.D., Duke-NUS Graduate Medical School Singapore;  
Duke University Medical Center, Durham, NC

### Abstract

In this chapter we discuss common challenges in and principles for conducting systematic reviews of genetic tests. The types of genetic tests discussed are those used to (1) determine risk or susceptibility in asymptomatic individuals; (2) reveal prognostic information to guide clinical management in those with a condition; or (3) predict response to treatments or environmental factors. This chapter is not intended to provide comprehensive guidance on evaluating all genetic tests. Rather, it focuses on issues that have been of particular concern to analysts and stakeholders and on areas that are of particular relevance for the evaluation of studies of genetic tests.

### Introduction

With recent advances in genotyping, it is expected that whole genome sequencing will soon be available for less than \$1,000. Consequently, the number of studies of genetic tests will likely increase substantially, as will the need to evaluate studies of genetic tests. The general principles for evaluating genetic tests are similar to those for interpreting other prognostic or predictive tests, but there are differences in how the principles need to be applied and the degree to which certain issues are relevant, particularly when considering genetic test results that provide predictive rather than diagnostic information.

This chapter focuses on issues of particular concern to analysts and stakeholders and areas of particular relevance for the evaluation of studies of genetic tests. It is not intended to provide comprehensive guidance on evaluating all genetic tests. We reflect on genetic tests used to (1) determine risk or susceptibility in asymptomatic individuals (e.g., to identify individuals at risk for future health conditions, such as BRCA1 and BRCA2 for breast and ovarian cancer); (2) reveal prognostic information to guide clinical management and treatment in those with a condition (e.g., Oncotype Dx<sup>®</sup> for breast cancer recurrence, a test to evaluate the tumor genome

of surgically excised tumors from patients with breast cancer); or (3) predict response to treatments or environmental factors including diet (nutrigenomics), drugs (pharmacogenomics, such as CYP2C9 and VKORC1 tests to inform warfarin dosing), infectious agents, chemicals, physical agents, and behavioral factors. We do not address genetic tests used for diagnostic purposes. We address issues related to both heritable mutations and somatic mutations (e.g., genetic tests for tumors).

Clinicians, geneticists, analysts, policymakers, and other stakeholders may have varying definitions of what is considered a “genetic test.” We have chosen to use a broad definition in agreement with that of the Centers for Disease Control and Prevention (CDC)–sponsored Evaluation of Genomic Applications in Practice and Prevention (EGAPP) and the Secretary’s Advisory Committee on Genetics, Health, and Society,<sup>1</sup> namely: “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.”<sup>1</sup> The same technologies are used for diagnostic and predictive genetic tests; it is the intended use of the test result that determines whether it is a diagnostic or predictive test.

In this chapter we discuss principles for addressing challenges related to developing the topic and structuring a genetic test review (context and scoping), as well as performing the review. This chapter is meant to complement the Methods Guide for Comparative Effectiveness Reviews.<sup>2</sup> We do not attempt to reiterate the challenges and principles described in earlier sections of this *Medical Test Methods Guide*, but focus instead on issues of particular relevance for evaluating studies of genetic tests. Although we have written this chapter to serve as guidance for the Agency for Healthcare Research and Quality (AHRQ) Evidence-based Practice Centers (EPCs), we also intend it as a useful resource for other investigators interested in conducting systematic reviews on genetic tests.

## Common Challenges

Genetic tests are different from other medical tests in their relationship to the outcomes measured. Reviewers need to take into account the penetrance of the disease, time lag to outcomes, variable expressivity, and pleiotropy (as defined below). These particular aspects of genetic tests result in specific actions at various stages of planning and performing the review. Both single-gene and polygenic disorders are known. Single gene disorders are the result of a single mutated gene and may be passed on to subsequent generations in various well described ways (e.g., autosomal dominant, autosomal recessive, X-linked). Polygenic disorders are the result of the combined action of more than one gene and are not inherited according to simple Mendelian patterns. Some examples include heart disease and diabetes. Some of the terms described below (penetrance, variable expressivity, and pleiotropy) are generally used to describe single-gene disorders.

## Penetrance

Evaluations of predictive genetic tests should always consider penetrance, defined as “the proportion of people with a particular genetic change who exhibit signs and symptoms of a disorder.”<sup>3</sup> Penetrance is a key factor in determining the future risk of developing disease and assessing the overall clinical utility of predictive genetic tests. Sufficient data to determine precise estimates of penetrance are sometimes lacking.<sup>4-5</sup> This can be due to a lack of reliable prevalence data or of long-term outcomes data. In such cases, determining the overall clinical

utility of a genetic test is difficult. In some cases, modeling with sensitivity analyses can help develop estimates.<sup>4</sup>

## **Time Lag**

The time lag between genetic testing and clinically important events should be assessed in critical appraisal of studies of such tests. Whether the duration of studies is sufficient to characterize the relationship between positive tests and clinical outcomes is an important consideration. In addition, it should be determined whether or not subjects have reached the age beyond which clinical expression is likely.

## **Variable Expressivity**

Variable expressivity refers to the range of severity of the signs and symptoms that can occur in different people with the same condition.<sup>3</sup> For example, the features of hemochromatosis vary widely. Some individuals have mild symptoms, while others experience life-threatening complications such as liver failure. The degree of expressivity should be considered in the evaluation of genetic tests.

## **Pleiotropy**

Pleiotropy occurs when a single gene influences multiple phenotypic traits. For example, the genetic mutation causing Marfan syndrome results in cardiovascular, skeletal, and ophthalmologic abnormalities. Similarly, BRCA mutations can increase the risk of a number of cancers, including breast, ovarian, prostate, and melanoma.

## **Other Common Challenges**

Another common challenge for the evaluation of predictive genetic tests is that direct evidence is often lacking on the impact of the test results on health outcomes. The evidence base is often too limited in scope to evaluate the clinical utility of the test. In addition, it is often difficult to find published information on various aspects of genetic tests, especially data related to analytic validity. For example, laboratory-developed tests (LDT) are regulated by the Centers for Medicare & Medicaid Services (CMS) under Clinical Laboratory Improvement Act (CLIA) regulations for clinical laboratories. CLIA does not require clinical validation and many LDTs have had no clinical validation or clinical utility studies.

Genetic tests also raise a number of technical issues particularly relevant to the assessment of their analytic validity. These technical issues may differ according to the type of genetic test and may influence the interpretation of a genetic test result. Technical issues may also differ depending on the specimen being tested. For example, there are different considerations when assessing tumor genomes as opposed to human genomes.

Several common challenges arise in using genetic tests to determine susceptibility or risk in asymptomatic individuals. The utility of such tests may depend on the ability of respondents, such as the patient or their relative, to report and identify certain clinical factors. For instance, if patients cannot accurately recall the family history of a heritable disease, it can be difficult to assess their risk of developing the disease.

Finally, statistical issues must be taken into account when evaluating studies of genetic tests. For example, genetic test results are often derived from analytically complex studies that have undergone a very large number of statistical tests, creating a high risk of Type I error (i.e., when a spurious association is deemed significant).

## Principles for Addressing the Challenges

The eight principles described in this section can be used to address the challenges related to developing, structuring, and performing a genetic test review (Table 11–1).

**Table 11–1. Principles for addressing common challenges when evaluating genetic tests used as predictive indicators**

<p><b>Principle 1:</b> Use an organizing framework appropriate for genetic tests.</p> <p><b>Principle 2:</b> Develop analytic frameworks that reflect the predictive nature of genetic tests and incorporate appropriate outcomes.</p> <p><b>Principle 3:</b> Search databases appropriate for genetic tests.</p> <p><b>Principle 4:</b> Consult with experts to determine which technical issues are important to address in assessing genetic tests.</p> <p><b>Principle 5:</b> Distinguish between functional assays and DNA-based assays to determine important technical issues.</p> <p><b>Principle 6:</b> Evaluate case-control studies carefully for potential selection bias.</p> <p><b>Principle 7:</b> Determine the added value of the genetic test over existing risk assessment approaches.</p> <p><b>Principle 8:</b> Understand statistical issues of particular relevance to genetic tests.</p>
--

### Principle 1: Use an organizing framework appropriate for genetic tests.

Organizing frameworks for evaluating genetic tests have been developed by the United States Preventive Services Task Force (USPSTF), the CDC, and EGAPP.<sup>1,6–7</sup> The model endorsed by the EGAPP initiative<sup>1</sup> was based on a previous report of the NIH Task Force on Genetic Testing<sup>8</sup> and developed through a CDC-sponsored project, which piloted an evidence evaluation framework that applied the following three criteria: (1) analytic validity (technical accuracy and reliability); (2) clinical validity (ability to detect or predict an outcome, disorder, or phenotype); and (3) clinical utility (whether use of the test to direct clinical management improves patient outcomes). A fourth criterion was added: (4) ethical, legal, and social implications.<sup>6</sup> The ACCE model (Analystic validity, Clinical validity, Clinical utility, and Ethical, legal and social implications) includes a series of 44 questions that are useful for analysts in defining the scope of a review, as well as for critically appraising studies of genetic tests (Table 11-2). The initial seven questions help to guide an understanding of the disorder, the setting, and the type of testing. A detailed description of the methods of the EGAPP Working Group is published elsewhere.<sup>1</sup>

### Principle 2: Develop analytic frameworks that reflect the predictive nature of genetic tests and incorporate appropriate outcomes.

It is important to have a clear definition of the clinical scenario and analytic framework when evaluating any test, including a predictive genetic test. Prior to performing a review, analysts should develop clearly defined key questions and understand the needs of decisionmakers and the context in which the tests are used. They should consider whether this is a test used for determining future risk of disease in asymptomatic individuals, establishing prognostic information that will influence treatment decisions, or predicting response to treatments (either effectiveness or harms)—or whether it is used for some other purpose. They should clarify the type of specimens used for the genetic test under evaluation (i.e., patient genome or tumor genome). The PICOTS typology (Patient population, Intervention, Comparator, Outcomes, Timing, Setting) should be clearly described, as it will inform the development of the analytic framework and vice versa.

**Table 11–2. ACCE model questions for reviews of genetic tests<sup>6</sup>**

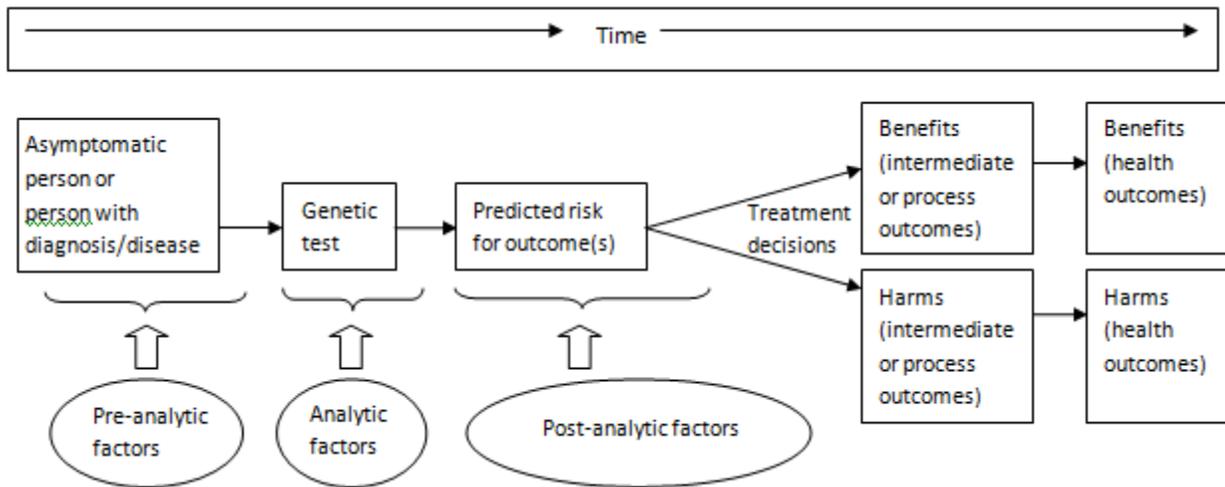
Element	Questions
Disorder/setting	1. What is the specific clinical disorder to be studied? 2. What are the clinical findings defining this disorder? 3. What is the clinical setting in which the test is to be performed? 4. What DNA test(s) are associated with this disorder? 5. Are preliminary screening questions employed? 6. Is it a stand-alone test or is it one of a series of tests? 7. If it is part of a series of screening tests, are all tests performed in all instances (parallel) or are only some tests performed on the basis of other results (series)?
Analytic validity	8. Is the test qualitative or quantitative? 9. How often is the test positive when a mutation is present? <sup>*</sup> 10. How often is the test negative when a mutation is not present? <sup>*</sup> 11. Is an internal quality control program defined and externally monitored? 12. Have repeated measurements been made on specimens? 13. What is the within-and between-laboratory precision? 14. If appropriate, how is confirmatory testing performed to resolve false positive results in a timely manner? 15. What range of patient specimens have been tested? 16. How often does the test fail to give a usable result? 17. How similar are results obtained in multiple laboratories using the same, or different technology?
Clinical validity	18. How often is the test positive when the disorder is present? 19. How often is the test negative when a disorder is not present? 20. Are there methods to resolve clinical false positive results in a timely manner? 21. What is the prevalence of the disorder in this setting? 22. Has the test been adequately validated on all populations to which it may be offered? 23. What are the positive and negative predictive values? 24. What are the genotype/phenotype relationships? <sup>*</sup> 25. What are the genetic, environmental or other modifiers? <sup>*</sup>
Clinical utility	26. What is the natural history of the disorder? 27. What is the impact of a positive (or negative) test on patient care? 28. If applicable, are medical tests available? 29. Is there an effective remedy, acceptable action, or other measurable benefit? 30. Is there general access to that remedy or action? 31. Is the test being offered to a socially vulnerable population? 32. What quality assurance measures are in place? 33. What are the results of pilot trials? 34. What health risks can be identified for follow-up testing and/or intervention? 35. What are the financial costs associated with testing? 36. What are the economic benefits associated with actions resulting from testing? 37. What facilities/personnel are available or easily put in place? 38. What educational materials have been developed and validated and which of these are available? <sup>*</sup> 39. Are there informed consent requirements? <sup>*</sup> 40. What methods exist for long term monitoring? 41. What guidelines have been developed for evaluating program performance?
Ethical, legal, and social implications	42. What is known about stigmatization, discrimination, privacy/confidentiality and personal/family social issues? <sup>*</sup> 43. Are there legal issues regarding consent, ownership of data and/or samples, patents, licensing, proprietary testing, obligation to disclose, or reporting requirements? <sup>*</sup> 44. What safeguards have been described and are these safeguards in place and effective? <sup>*</sup>

ACCE = Analytic validity, Clinical validity, Clinical utility, and Ethical, legal and social implications;  
 DNA = deoxyribonucleic acid

<sup>\*</sup>Many of the questions in this Table (or variants of the questions) are relevant for evaluating most medical tests, not just genetic tests. Those with an asterisk (questions 4, 9, 10, 24, 25, 38, 39, 42, 43, and 44) are relevant only for evaluating genetic tests or may require extra scrutiny when evaluating genetic tests.

In constructing an analytic framework, it may be useful for analysts to consider preanalytic, analytic, and postanalytic factors particularly applicable to genetic tests (described later in this chapter), as well as the key outcomes of interest. Analytic frameworks should incorporate the factors and outcomes of greatest interest to decision makers. Figure 11–1 illustrates a generic analytic framework for evaluating predictive genetic tests that can be modified as necessary for various situations.

**Figure 11–1. Generic analytic framework for evaluating predictive genetic tests**

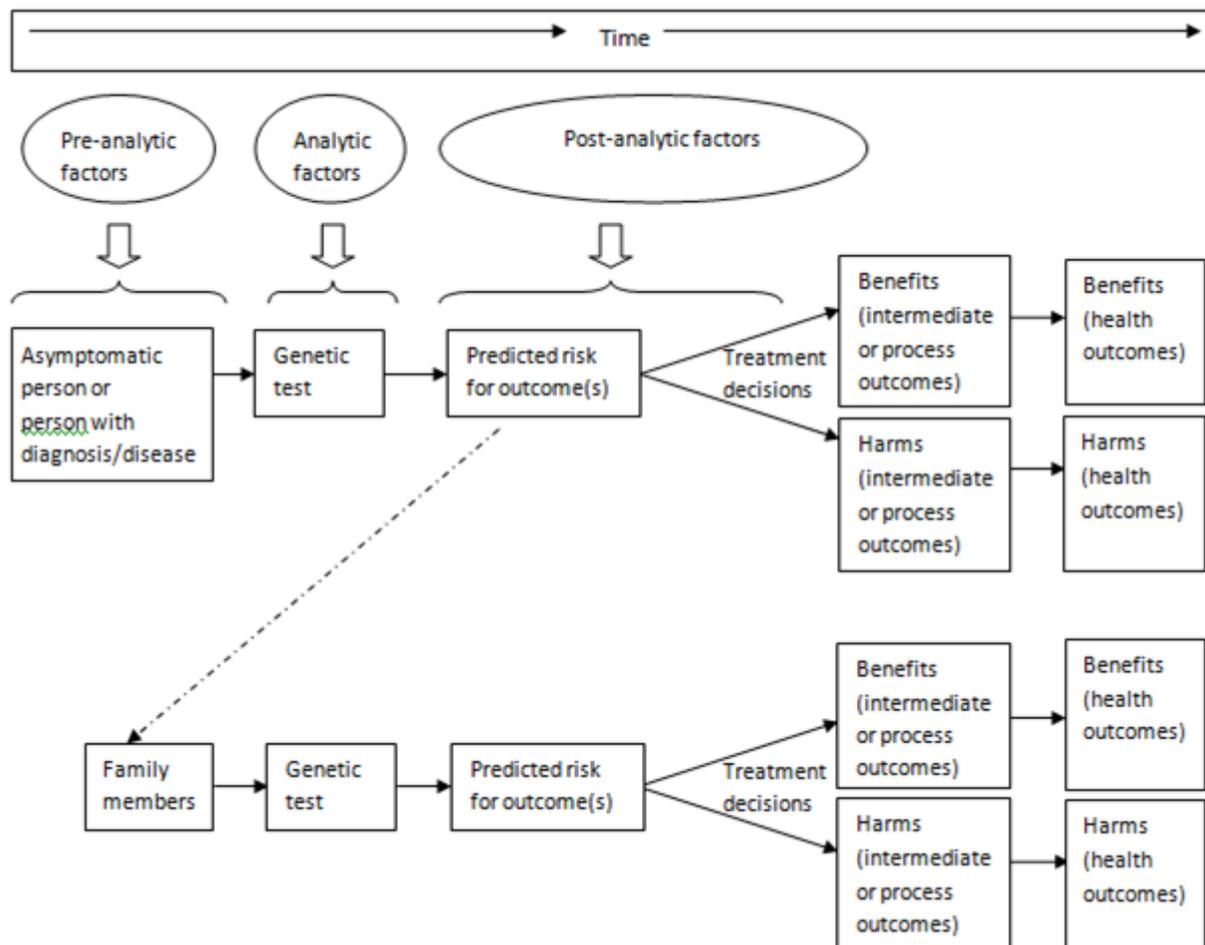


In addition to effects on family members, psychological distress and possible stigmatization or discrimination are potential harms that may result from predictive genetic tests, particularly if the test results predict probability of disease occurring with a high likelihood, especially if no proven preventive or ameliorative measures are available. For these potential harms, analysts should take into account whether the testing is for inherited or acquired genetic mutations, since these factors influence the potential for harms. In addition, whether the condition related to the test is multifactorial or follows classic Mendelian inheritance will affect the potential for these harms.

Other important outcomes to consider when evaluating genetic tests include, but are not limited to, cost, quality of life, long-term morbidity, and indirect impact. Genetic tests may have an impact that is difficult to measure, such as impact on important decisions regarding pregnancy.

Depending on the context, the impact of genetic testing on family members may be important, particularly in cases that involve testing for heritable conditions. One approach to including family members in the analytic framework is illustrated in Figure 11–2.

**Figure 11–2. Generic analytic framework for evaluating predictive genetic tests when the impact on family members is important**



### **Principle 3: Search databases appropriate for genetic tests.**

The Human Genome Epidemiology Network (HuGE Net) Web site can provide a helpful supplement to searches, as it includes many meta-analyses of genetic association studies as well as a source called the HuGE Navigator that can identify all types of available studies related to a genetic test.<sup>9</sup>

The U.S. Food and Drug Administration (FDA)–approved test package inserts for genetic tests contain summaries of the analytic validity data. These summaries can be retrieved through searches of the gray literature. Package inserts are available on the FDA and manufacturer Web sites. Laboratory-developed tests do not require FDA clearance, and there is no requirement for publicly available data on analytic validity. When there are no published data on analytic validity of a genetic test, the external proficiency testing program carried out jointly by the American College of Medical Genetics (ACMG) and the College of American Pathologists (CAP) can be useful in establishing the degree of laboratory-to-laboratory variability, as well as some sense of reproducibility.<sup>10–12</sup> Other potentially useful sources of unpublished data include conference publications from professional societies (e.g., the College of American Pathologists), the GeneTests Web site ([www.genetests.org](http://www.genetests.org)), the Association for Molecular Pathology Web site ([www.amp.org](http://www.amp.org)), CDC programs (e.g., the Genetic Testing Reference Materials Coordination

Program and the Newborn Screening Quality Assurance Program), and international proficiency testing programs.<sup>13</sup>

An AHRQ “horizon scan” found two databases—LexisNexis® (www.lexisnexis.com) and Cambridge Healthtech Institute (CHI) (www.healthtech.com/)—that had high utility in identifying genetic tests in development for clinical cancer care. A number of others had low-to-moderate utility, and some were not useful.<sup>14</sup>

### **Principle 4: Consult with experts to determine which technical issues are important to address in assessing genetic tests.**

There are a number of technical issues related to analytic validity that can influence the interpretation of a genetic test result, including preanalytic, analytic, and postanalytic factors.<sup>15–16</sup> In general, preanalytic steps are those involved in obtaining, fixing or preserving, and storing samples prior to staining and analysis. Important analytic variables include the type of assay chosen and its reliability, types of samples, the specific analyte investigated, specific genotyping methods, timing of sample analysis, and complexity of performing the assay. Postanalytic variables relate to the complexity of interpreting the test result, variability from laboratory to laboratory, and quality control.<sup>15–16</sup> To determine which of these technical issues are pertinent for a given review, comparative effectiveness review teams should include or consult with molecular pathologists, geneticists, or others familiar with the issues related to the process of performing and reporting genetic tests. Table 11–3 summarizes some of the preanalytic, analytic, and postanalytic questions that should be addressed.

For genetic testing of tumor specimens, it is important to understand that the tumor genome may be in a dynamic state, with mutations emerging over time (e.g., due to drug exposure or disruption of cellular repair). Tumor specimens will often contain normal cells from the patient as well as tumor cells. To accurately assess for somatic mutations using tumor specimens, particular strategies may be needed, such as enriching samples for tumor cells (e.g., by microscopic evaluation and dissection of the cells).

**Table 11–3. Questions for assessing preanalytic, analytic, and postanalytic factors for evaluating predictive genetic tests\***

Element	Questions
Preanalytic	What patient characteristics are relevant to the analytic validity of the test (e.g., age, sex, ethnicity, race, ancestry, parental history of consanguinity, family health history)? What types of samples were used? How were samples obtained? How were samples handled and stored prior to analysis?
Analytic	What type of assay was used? What is the reliability of the assay? What specific analyte was investigated (e.g., specification of which alleles, genes, or biochemical analytes were evaluated)? For DNA-based tests, what is the definition of the genotype investigated? Did the study test for all potentially relevant alleles? For DNA-based tests, what genotyping methods were used? When were samples analyzed (compared to when they were collected)? Was the timing of analysis equal for both study groups (if applicable)? How often does the test give a usable result? (i.e., What is the “call rate”?)
Postanalytic	How are the test results interpreted and applied? How complex is interpretation and application? What quality control measures were used? Were repeated measurements made on specimens? How reproducible is the test over time? How reproducible is the test when repeated in the same patient multiple times? How reproducible is the test from laboratory to laboratory?

\*Portions of this table were adapted from Burke et al., 2002<sup>15</sup> and Little et al., 2002.<sup>16</sup>

## **Principle 5: Distinguish between functional assays and DNA-based assays to determine important technical issues.**

Some studies may utilize DNA-based assays, whereas others may utilize functional assays with different sensitivities and specificities. Functional assays, in which a substrate or product of a metabolic process affected by a particular genetic polymorphism is measured, may have the advantage of showing potentially more important information than the presence of the genetic polymorphism itself. However, they may be affected by a number of factors and do not necessarily reflect the polymorphism alone. Unmeasured environmental factors, other genetic polymorphisms, and various disease states may influence the results of functional assays. In addition, functional assays that measure enzyme activity are taken at a single point in time. Depending on the enzyme and polymorphism being evaluated, the variation in enzyme activity over time should be considered in critical appraisal. Inconsistent results have been reported between studies using DNA-based molecular methods and those using phenotypic assays.<sup>16-18</sup>

For DNA-based tests, a variety of sample sources are available (e.g., blood, cheek swab, hair) that should hypothetically result in identical genotype results.<sup>16,19-23</sup> However, DNA may be more difficult to obtain and purify from some tissues than from blood, particularly if the tissues have been fixed in paraffin versus fresh samples. (DNA extraction from formalin-fixed tissue is difficult, but sometimes possible).<sup>16</sup> Some studies utilize different sources of DNA for cases and controls, introducing potential measurement bias from differences in ease of technique and test accuracy. Extraction of DNA from tumors in oncology studies may raise additional issues that influence analytic validity, including the quantity of tissue, admixture of normal and cancerous tissue, amount of necrosis, timing of collection, and storage technique (e.g., fresh, frozen, paraffin, formalin).<sup>16</sup>

When evaluating DNA-based molecular tests, complexity of the test method, laboratory-to-laboratory variability, and quality control should be assessed. A number of methods are available for genotyping single nucleotide polymorphisms that vary in complexity and potential for polymorphism misclassification.<sup>16,24-26</sup> Considering laboratory reporting of internal controls and repetitive experiments can be useful in assessment of overall analytic validity. The method of interpreting test results may influence complexity as well. For example, some tests require visual inspection of electrophoresis gels. Inter-observer variability should be considered for such tests.<sup>16,27</sup>

## **Principle 6: Evaluate case-control studies carefully for potential selection bias.**

In critical appraisal of any case-control study, it is important to determine whether cases and controls were selected from the same source population. In the case of genetic studies, the geographic location of the population does not suffice. Rather, having cases and controls matched for ethnicity/race or ancestry (i.e., population stratification) is important, since the frequencies of DNA polymorphisms vary from population to population. It has been noted that many case-control studies of gene-disease associations have selected controls from a population that does not represent the population from which the cases arose.<sup>16-17,28-30</sup> In general, only nested case-control studies could have low enough potential for selection bias to provide reliable information.

## **Principle 7: Determine the added value of the genetic test over existing risk assessment approaches.**

For some scenarios, a number of clinical factors associated with risk assessment or susceptibility may already be well characterized. In such cases, comparative effectiveness reviews should determine the added value of using genetic testing along with known factors, compared with using the known factors alone. For example, age, sex, smoking, hypertension, diabetes, and cholesterol are all well established risk factors for cardiovascular disease. Risk stratification of individuals to determine cholesterol-lowering targets is based on these factors.<sup>31</sup> Assessment of newly identified polymorphisms—such as those described on chromosome 9p21<sup>32</sup>—that may confer increased risk of cardiovascular disease and have potential implications for medical interventions should be evaluated in the context of these known risk factors. In this scenario, investigators should determine the added value of testing for polymorphisms of chromosome 9p21 in addition to known clinical risk factors.

Multiple polymorphisms may be associated with risk of disease, prognosis, or prediction of drug response. In such cases, the effect of multiple polymorphisms can be explored using a multiple regression model. Once this is done, prospective studies would usually be needed to determine whether the model, including the genetic tests, has clinical utility. For example, VKORC1 and CYP2C9 genotypes have been associated with warfarin dose requirements in multiple regression models. In order to determine whether tests for VKORC1 and CYP2C9 have clinical utility, studies would need to compare the use of a prediction model that contains the genetic tests in combination with known clinical factors that affect warfarin dose (e.g., age, BMI) with the use of clinical factors alone.<sup>33–35</sup>

## **Principle 8: Understand statistical issues of particular relevance to genetic tests.**

### **Hardy-Weinberg Equilibrium**

In population genetics, most allele distributions follow a usual distribution, known as the Hardy-Weinberg equilibrium (HWE). Genetic association studies should generally report whether the frequencies of the alleles being evaluated follow HWE. There are a number of reasons that distributions may deviate from HWE, including new mutations, selection, migration, genetic drift, and inbreeding.<sup>36</sup> In addition, when numerous polymorphisms are tested for associations with diseases or outcomes, as in many genome-wide association studies, many of them (5 percent) will deviate from HWE based on chance alone (related to multiple testing).<sup>37</sup> Deviation from HWE may be a clue to bias and genotyping error, but it is not specific and possibly not sensitive.<sup>37</sup> Analysts should consider whether studies have tested for and reported HWE. A more detailed discussion of this topic as it relates to genetic association studies has been published elsewhere.<sup>36–37</sup>

### **Sample Size Calculations**

When assessing internal validity of studies, it is important to assess whether sample size calculations appropriately accounted for the number of variant alleles and the prevalence of variants in the population of interest. This is particularly relevant for pharmacogenomic studies evaluating the functional relevance of genetic polymorphisms.<sup>38</sup> Such studies often enroll an insufficient number of subjects to account for the number of variant alleles and the prevalence of variants in the population.<sup>38</sup>

## Genetic Association Studies and Multiple Comparisons

Genetic test results are sometimes derived from analytically complex studies that have undergone a very large number of statistical tests. These may be in the form of genome-wide association studies searching for associations between a huge number of genetic polymorphisms and health conditions. Such association studies may enhance understanding of the importance of genetics in relation to a variety of health conditions, but should generally be used to generate hypotheses rather than to test hypotheses or to confirm cause-effect relationships.<sup>16</sup> Close scrutiny should be applied to ensure that the evidence for the association has been validated in multiple studies to minimize both potential confounding and potential publication bias issues. In addition, reviewers should note whether appropriate adjustments for multiple comparisons were used. Many investigators recommend using a P value of less than  $5 \times 10^{-8}$  for the threshold of significance in large genome-wide studies.<sup>37,39-40</sup> Other approaches include assessing the false positive report probability and controlling the false discovery rate.<sup>41-43</sup>

When a genetic mutation associated with increased risk is present, evaluating potential causality can be difficult, as many factors other than the mutation may influence associations. These include environmental exposures, behaviors, and other genes. Many genetic variants identified that are thought to influence susceptibility to diseases are associated with low relative and absolute risk.<sup>16,44</sup> Thus, exclusion of non-causal explanations for associations and consideration of potential confounders are central to critical appraisal of such associations. It may also be important to explore biologic plausibility (e.g., from *in vitro* studies) to help support or oppose theories of causation.<sup>16</sup>

## Overlapping Data Sets

Be cautious of publications that report prevalence estimates for genetic variants that have actually arisen from overlapping data sets.<sup>16</sup> For example, genome-wide association studies or other large collaborative efforts, such as the International Warfarin Pharmacogenomics Consortium, may pool samples of patients that were previously included in other published studies.<sup>3</sup> To the degree possible, investigators should identify overlapping data sets and avoid double-counting. It may be useful to organize evidence tables by study time period and geographic area to identify potential overlapping data sets.<sup>16</sup>

## Assessing Tumor Genetics

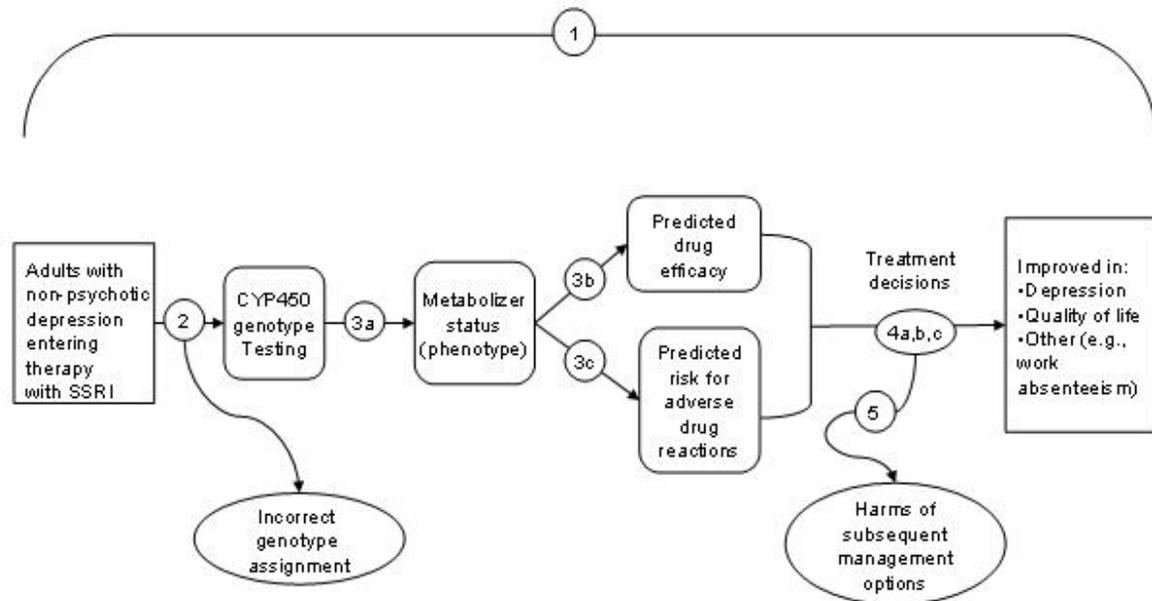
As mentioned under Principle 4, it is important to understand that a tumor genome may be in a dynamic state. In addition, tumor specimens will often contain normal cells from the patient. The characteristics of the specimen will influence the sensitivity and operating characteristics of the test. Tests with greater sensitivity may be required when specimens contain both normal cells and tumor cells.

## Illustrations

Since the completion of the Human Genome Project, the Hap Map project, and related works, there have been a great number of publications describing the clinical validity of genetic test results (e.g., gene-disease associations), but far fewer studies of their clinical utility. A review of genetic testing for cytochrome P450 polymorphisms in adults with depression treated with selective serotonin reuptake inhibitors (SSRIs) developed an analytic framework and five corresponding key questions which, taken together, provide an example of a well defined predictive genetic test scenario that explores a potential chain of evidence relating to intermediate outcomes (Figure 11-3).<sup>45</sup> The authors found no prospective studies with clinical outcomes that used genotyping to guide treatment. They constructed a chain of questions to

assess whether sufficient indirect evidence could answer the overarching question by evaluating the links between genotype and metabolism of SSRIs (phenotype), metabolism and SSRI efficacy, and metabolism and adverse drug reactions to SSRIs.

**Figure 11–3. Analytic framework for evidence gathering on CYP450 genotype testing for SSRI treatment of depression.**



CYP450 = cytochrome p450; SSRI = selective serotonin reuptake inhibitor

Numbers in this figure represent the research questions addressed in the systematic review:<sup>45</sup>

**1 (overarching question):** Does testing for cytochrome P450 (CYP450) polymorphisms in adults entering selective serotonin reuptake inhibitor (SSRI) treatment for non-psychotic depression lead to improvement in outcomes, or are testing results useful in medical, personal, or public health decisionmaking?

**2:** What is the analytic validity of tests that identify key CYP450 polymorphisms?

**3a:** How well do particular CYP450 genotypes predict metabolism of particular SSRIs? Do factors such as race/ethnicity, diet, or other medications, affect this association?

**3b:** How well does CYP450 testing predict drug efficacy? Do factors such as race/ethnicity, diet, or other medications, affect this association?

**3c:** How well does CYP450 testing predict adverse drug reactions? Do factors such as race/ethnicity, diet, or other medications, affect this association?

**4a:** Does CYP450 testing influence depression management decisions by patients and providers in ways that could improve or worsen outcomes?

**4b:** Does the identification of the CYP450 genotypes in adults entering SSRI treatment for non-psychotic depression lead to improved clinical outcomes compared to not testing?

**4c:** Are the testing results useful in medical, personal or public health decisionmaking?

**5:** What are the harms associated with testing for CYP450 polymorphisms and subsequent management options?

An EPC report on HER2 testing to manage patients with breast cancer and other solid tumors provides a detailed assessment of challenges in conducting a definitive evaluation of preanalytic, analytic, and postanalytic factors when there is substantial heterogeneity or lack of available information related to the methods of testing.<sup>46</sup> The authors noted that it had been only very recently that many aspects of HER2 assays were standardized, and that the effects of widely varying testing methods could not be isolated. Thus, they approached this challenge by providing a narrative review for their first key question (What is the evidence on concordance and discrepancy rates for methods [e.g., FISH, IHC, etc.] used to analyze HER2 status in breast tumor tissue?).

Additional considerations arise when evaluating genetic test results used to determine susceptibility or risk in asymptomatic individuals. The utility of such tests may depend on the ability of patients and providers to report and identify certain clinical factors. For example, a review of genetic risk assessment and BRCA mutation testing underscores the importance of accurately determining family history.<sup>4,47</sup> The analytic framework begins by classifying asymptomatic women into high, moderate, or average risk categories. This is a good example of incorporating a key preanalytic factor (family history), that has an important influence on analytic validity. Tests for BRCA mutations may be used to predict the risk for breast and ovarian cancer in high-risk women (i.e., those with a family history suggesting increased risk). However, because we do not know all of the genes that contribute to hereditary breast and ovarian cancer and because analytic methods to detect mutations in the known genes are not perfect, population-based testing for hereditary susceptibility to breast and ovarian cancer is currently not an appropriate strategy. Rather, family history-based testing is the paradigm that is recommended to guide the use of BRCA testing.<sup>4,47</sup>

Thus, family history is a genetic/genomics tool that is used to (1) identify people with possible inherited disease susceptibilities, (2) guide genetic testing strategies, (3) help interpret genetic test results, and (4) assess disease risk. The ability of providers to accurately determine a family history that confers increased risk is a key prerequisite to the utility of BRCA mutation and other predictive genetic testing. It is sometimes difficult for people to accurately recall the presence of a condition in their relatives. Sensitivity and specificity of self-reported family history are important in determining overall usefulness of predictive genetic testing.<sup>4</sup>

## Conclusions

Analysts should understand common challenges, and apply the principles for addressing those challenges, when conducting systematic reviews of genetic tests used as predictive indicators. Key points include:

1. The general principles that apply in evaluating genetic tests are similar to those for other prognostic or predictive tests, but there are differences in how the principles need to be applied or the degree to which certain issues are relevant.
2. A clear definition of the clinical scenario and an analytic framework is important when evaluating *any* test, including genetic tests.
3. Organizing frameworks and analytic frameworks are useful constructs for approaching the evaluation of genetic tests.
4. In constructing an analytic framework for evaluating a genetic test, analysts should consider preanalytic, analytic, and postanalytic factors; such factors are useful when assessing analytic validity.
5. Predictive genetic tests are generally characterized by a delayed time between testing and clinically important events.
6. Published information on the analytic validity of some genetic tests may be difficult to find. Web sites (FDA or diagnostic companies) and gray literature may be important sources.
7. In situations where clinical factors associated with risk are well characterized, comparative effectiveness reviews should assess the added value of using genetic testing along with known factors, compared with using the known factors alone.
8. For genome-wide association studies, reviewers should determine whether the association has been validated in multiple studies to minimize both potential confounding and publication bias. In addition, reviewers should note whether appropriate adjustments for multiple comparisons were used.

## References

1. 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*. 2008.
2. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. AHRQ Publication No. 10(11)-EHC063-EF. Rockville, MD: Agency for Healthcare Research and Quality; March 2011; Available at: [www.effectivehealthcare.ahrq.gov](http://www.effectivehealthcare.ahrq.gov). Accessed August 22, 2011.
3. Lister Hill National Center for Biomedical Communications: Collections of the National Library of Medicine. What are reduced penetrance and variable expressivity? [electronic resource]; 2008; Available at: <http://ghr.nlm.nih.gov/handbook/inheritance/pentranceexpressivity>. Accessed August 22, 2011.
4. Nelson HD, Huffman LH, Fu R, Harris EL. Genetic risk assessment and BRCA mutation testing for breast and ovarian cancer susceptibility: systematic evidence review for the U.S. Preventive Services Task Force. *Annals of internal medicine*. 2005;143(5):362-79.
5. 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. *Annals of internal medicine*. 2006;145(3):209-23.
6. National Office of Public Health Genomics C. ACCE Model Process for Evaluating Genetic Tests. 2007; Available at: <http://www.cdc.gov/genomics/gtesting/ACCE/index.htm>. Accessed August 22, 2011.
7. Harris RP, Helfand M, Woolf SH, et al. Current methods of the US Preventive Services Task Force: a review of the process. *American journal of preventive medicine*. 2001;20(3 Suppl):21-35.
8. Task Force on Genetic Testing (NIH). Promoting Safe and Effective Genetic Testing in the United States. Final Report of the Task Force on Genetic Testing. 1997; Available at: <http://www.genome.gov/10001733>. Accessed August 22, 2011.
9. Khoury MJ, Dorman JS. The Human Genome Epidemiology Network. *American journal of epidemiology*. 1998;148(1):1-3.
10. Palomaki GE, Bradley LA, Richards CS, Haddow JE. Analytic validity of cystic fibrosis testing: a preliminary estimate. *Genet Med*. 2003;5(1):15-20.
11. Palomaki GE, Haddow JE, Bradley LA, FitzSimmons SC. Updated assessment of cystic fibrosis mutation frequencies in non-Hispanic Caucasians. *Genet Med*. 2002;4(2):90-4.
12. Palomaki GE, Haddow JE, Bradley LA, 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(6):440-3.
13. Sun F, Bruening W, Erinoff E, Schoelles KM. Addressing Challenges in Genetic Test Evaluation. Evaluation Frameworks and Assessment of Analytic Validity. Methods Research Report (Prepared by the ECRI Institute Evidence-based Practice Center under Contract No. 290-2007-10063-I.) AHRQ Publication No. 11-EHC048-EF. Rockville, MD: Agency for Healthcare Research and Quality; June 2011. Available at: [www.effectivehealthcare.ahrq.gov/reports/final.cfm](http://www.effectivehealthcare.ahrq.gov/reports/final.cfm).
14. Agency for Healthcare Research and Quality. *Genetic Tests for Cancer*. Technology Assessment. Rockville, MD: Agency for Healthcare Research and Quality; 2006; Available at: <http://archive.ahrq.gov/clinic/ta/gentests/>. Accessed August 22, 2011.
15. Burke W, Atkins D, Gwinn M, et al. Genetic test evaluation: information needs of clinicians, policy makers, and the public. *American journal of epidemiology*. 2002;156(4):311-8.
16. Little J, Bradley L, Bray MS, et al. Reporting, appraising, and integrating data on genotype prevalence and gene-disease associations. *American journal of epidemiology*. 2002;156(4):300-10.
17. Brockton N, Little J, Sharp L, Cotton SC. N-acetyltransferase polymorphisms and colorectal cancer: a HuGE review. *American journal of epidemiology*. 2000;151(9):846-61.
18. d'Errico A, Malats N, Vineis P, Boffetta P. Review of studies of selected metabolic polymorphisms and cancer. *IARC scientific publications*. 1999(148):323-93.

19. Yang M, Hendrie HC, Hall KS, Oluwole OS, Hodes ME, Sahota A. Improved procedure for eluting DNA from dried blood spots. *Clinical chemistry*. 1996;42(7):1115-6.
20. Gale KB, Ford AM, Repp R, et al. Backtracking leukemia to birth: identification of clonotypic gene fusion sequences in neonatal blood spots. *Proceedings of the National Academy of Sciences of the United States of America*. 1997;94(25):13950-4.
21. Walker AH, Najarian D, White DL, Jaffe JF, Kanetsky PA, Rebbeck TR. Collection of genomic DNA by buccal swabs for polymerase chain reaction-based biomarker assays. *Environmental health perspectives*. 1999;107(7):517-20.
22. Harty LC, Shields PG, Winn DM, Caporaso NE, Hayes RB. Self-collection of oral epithelial cell DNA under instruction from epidemiologic interviewers. *American journal of epidemiology*. 2000;151(2):199-205.
23. Garcia-Closas M, Egan KM, Abruzzo J, et al. Collection of genomic DNA from adults in epidemiological studies by buccal cytobrush and mouthwash. *Cancer Epidemiol Biomarkers Prev*. 2001;10(6):687-96.
24. Hixson JE, Vernier DT. Restriction isotyping of human apolipoprotein E by gene amplification and cleavage with HhaI. *Journal of lipid research*. 1990;31(3):545-8.
25. Tobe VO, Taylor SL, Nickerson DA. Single-well genotyping of diallelic sequence variations by a two-color ELISA-based oligonucleotide ligation assay. *Nucleic acids research*. 1996;24(19):3728-32.
26. Lee LG, Connell CR, Bloch W. Allelic discrimination by nick-translation PCR with fluorogenic probes. *Nucleic acids research*. 1993;21(16):3761-6.
27. Bogardus ST, Jr., Concato J, Feinstein AR. Clinical epidemiological quality in molecular genetic research: the need for methodological standards. *JAMA*. 1999;281(20):1919-26.
28. Botto LD, Yang Q. 5,10-Methylenetetrahydrofolate reductase gene variants and congenital anomalies: a HuGE review. *American journal of epidemiology*. 2000;151(9):862-77.
29. Dorman JS, Bunker CH. HLA-DQ locus of the human leukocyte antigen complex and type 1 diabetes mellitus: a HuGE review. *Epidemiologic reviews*. 2000;22(2):218-27.
30. Cotton SC, Sharp L, Little J, Brockton N. Glutathione S-transferase polymorphisms and colorectal cancer: a HuGE review. *American journal of epidemiology*. 2000;151(1):7-32.
31. National Cholesterol Education Program. Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report; 2002 Dec 17. Report No.: 1524-4539 (Electronic).
32. Schunkert H, Gotz A, Braund P, et al. Repeated replication and a prospective meta-analysis of the association between chromosome 9p21.3 and coronary artery disease. *Circulation*. 2008;117(13):1675-84.
33. Gage BF, Eby C, Milligan PE, Banet GA, Duncan JR, McLeod HL. Use of pharmacogenetics and clinical factors to predict the maintenance dose of warfarin. *Thromb Haemost*. 2004;91(1):87-94.
34. Gage BF, Lesko LJ. Pharmacogenetics of warfarin: regulatory, scientific, and clinical issues. *J Thromb Thrombolysis*. 2008;25(1):45-51.
35. Jonas DE, McLeod HL. Genetic and clinical factors relating to warfarin dosing. *Trends Pharmacol Sci*. 2009;30(7):375-86.
36. Attia J, Ioannidis JP, Thakkinstian A, et al. How to use an article about genetic association: A: Background concepts. *JAMA*. 2009;301(1):74-81.
37. Attia J, Ioannidis JP, Thakkinstian A, et al. How to use an article about genetic association: B: Are the results of the study valid? *JAMA*. 2009;301(2):191-7.
38. Williams JA, Johnson K, Paulauskis J, Cook J. So many studies, too few subjects: establishing functional relevance of genetic polymorphisms on pharmacokinetics. *Journal of clinical pharmacology*. 2006;46(3):258-64.
39. Hoggart CJ, Clark TG, De Iorio M, Whittaker JC, Balding DJ. Genome-wide significance for dense SNP and resequencing data. *Genetic epidemiology*. 2008;32(2):179-85.
40. McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature reviews*. 2008;9(5):356-69.

41. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. *Behav Brain Res.* 2001;125(1-2):279-84.
42. Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J Natl Cancer Inst.* 2004;96(6):434-42.
43. Ziegler A, Konig IR, Thompson JR. Biostatistical aspects of genome-wide association studies. *Biom J.* 2008;50(1):8-28.
44. Caporaso N. Selection of candidate genes for population studies. In: Vineis P, Malats N, Lang M, et al, editors. *Metabolic polymorphisms and susceptibility to cancer.* Lyon, France: IARC Monogr Eval Carcinog Risks Hum; 1999. p. 23-36.
45. 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). *Evid Rep Technol Assess (Full Rep).* 2007(146):1-77.
46. Seidenfeld J, Samson DJ, Rothenberg BM, Bonnell CJ, Ziegler KM, Aronson N. *HER2 Testing to Manage Patients With Breast Cancer or Other Solid Tumors/Technology Assessment No. 172.* Rockville, MD: (Prepared by Blue Cross and Blue Shield Association Technology Evaluation Center Evidence-based Practice Center, under Contract No. 290-02-0026.) 2008.
47. Genetic risk assessment: recommendation statement. Genetic risk assessment and BRCA mutation testing for breast and ovarian cancer susceptibility: recommendation statement. *Annals of internal medicine.* 2005;143(5):355-61.

**Acknowledgements:** We would like to thank Halle R. Amick (University of North Carolina, Cecil G. Sheps Center for Health Services Research) and Crystal M. Riley (Duke-NUS Graduate Medical School Singapore) for their assistance with preparation of this manuscript, insightful editing, and outstanding attention to detail. We deeply appreciate the considerable support, commitment, and contributions of Stephanie Chang, MD, MPH, the AHRQ Task Order Officer for this project and the Evidence-based Practice Center Director.

**Funding:** Funded by the Agency for Health Care Research and Quality (AHRQ) under the Effective Health Care Program.

**Disclaimer:** The findings and conclusions expressed here are those of the authors and do not necessarily represent the views of AHRQ. Therefore, no statement should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

**Public domain notice:** This document is in the public domain and may be used and reprinted without permission except those copyrighted materials that are clearly noted in the document. Further reproduction of those copyrighted materials is prohibited without the specific permission of copyright holders.

**Accessibility:** Persons using assistive technology may not be able to fully access information in this report. For assistance contact [EffectiveHealthCare@ahrq.hhs.gov](mailto:EffectiveHealthCare@ahrq.hhs.gov).

**Conflict of interest:** None of the authors has any affiliations or financial involvement that conflicts with the information presented in this chapter.

**Corresponding author:** Daniel E. Jonas, M.D., M.P.H., 5034 Old Clinic Building, CB #7110, Chapel Hill, NC 27599. Phone (919) 966-7102; Fax (919) 966-2274; email [daniel\\_jonas@med.unc.edu](mailto:daniel_jonas@med.unc.edu)

**Suggested citation:** Jonas DE, Wilt TJ, Taylor BC, Wilkins TM, Matchar DB. Challenges in and principles for conducting systematic reviews of genetic tests used as predictive indicators. AHRQ Publication No. 12-EHC083-EF. Chapter 11 of Methods Guide for Medical Test Reviews (AHRQ Publication No. 12-EHC017). Rockville, MD: Agency for Healthcare Research and Quality; June 2012. [www.effectivehealthcare.ahrq.gov/reports/final.cfm](http://www.effectivehealthcare.ahrq.gov/reports/final.cfm). Also published in a special supplement to the Journal of General Internal Medicine, July 2012.