review

Genetic tests and their evaluation: Can we answer the key questions?

Mark Kroese, MRCGP, MFPHM, Ron L. Zimmern, FRCP, FFPHM and Simon Sanderson, MRCP, MFPHM

The rapid pace of research in the field of genetics has already yielded many benefits. The development of new genetic tests is one such example. Before there can be widespread uptake of these tests they need to be evaluated to confirm the benefits of their use. The authors review some of the key features of the evaluation of diagnostic tests focusing on analytical and clinical validity. Test properties such as sensitivity, specificity, likelihood ratios, positive and negative predictive values, and how they relate to molecular genetic testing are discussed. Associated issues such as the concepts of disease definition, imperfect reference standards, and false positives are also explored. The authors suggest possible approaches to addressing some of the problems identified. ***Genet Med* 2004:6(6):475–480.**

Key Words: *evaluation, genetic test, test reference standards, test validity, predictive value*

The rapid pace of development in the field of genetics has increased our knowledge of the molecular basis of disease. This information is now being applied to the development of genetic tests, which it is hoped will enable clinicians to improve the care they provide for their patients. These efforts should be supported but it is essential that the tests are appropriately evaluated before widespread dissemination in clinical practice. There is evidence that diagnostic tests of all types, not just molecular genetic tests, are often implemented without adequate appraisal; the dexamethasone suppression test as a diagnostic aid for depression is one such example.[1–3] Diagnostic tests account for a significant proportion of health care budgets in the developed world; the potential availability of genetic tests will increase these numbers, and the benefits of further investment will need to be demonstrated.

We believe that genetic tests, by virtue of their use of DNA based technologies, give rise to few if any special or specific issues, but that, by focusing on their evaluation, issues that are relevant to all diagnostic technologies are raised. The transfer of tests developed in a research setting to clinical practice can present considerable difficulties. We present aspects of diagnostic test evaluation as it relates to molecular genetic testing and illustrate some of the problems that can be encountered. We also intend to show that a number of the issues are conceptual in nature, rather than technical or epidemiological. These are primarily about our understanding of the term "disease," one that may differ according to whether the disease in question is a single gene disorder or some other more complex entity.

TEST CHARACTERISTICS AND THEIR DEFINITIONS

A genetic test has been defined as the analysis of human DNA, RNA, chromosomes, proteins, and certain metabolites in order to detect heritable disease related genotypes, mutations, phenotypes, or karyotypes for clinical purposes.[4] There are many problems and issues that surround this definition; it will not be possible to address these in this review.[5] For our purposes, we focus on the concept of a gene test, defining that as one based on the analysis of human DNA using a variety of different technologies. In this review, any reference to a genetic test will be taken to refer to a gene test. We also note that a test may seek to predict or determine the susceptibility to, or probability of, developing disease in the future, as well as to establish a diagnosis in someone with clinical signs and symptoms.

Considerable work has already been performed to establish a framework for the assessment of genetic tests.[6–9] These have focused on four key areas: analytical validity, clinical validity, clinical utility and the ethical, legal and social implications. In this review, we confine our discussion to analytical and clinical validity; we shall discuss their use in both single gene disorders and complex disorders.

The analytical validity of a genetic test defines its ability to measure accurately and reliably the genotype of interest. Clinical validity defines its ability to detect or predict the presence or absence of the phenotype, physical trait, clinical disease, or predisposition to disease.[6] Measures of test performance are well described in textbooks of clinical epidemiology and are based around the concepts of sensitivity, specificity, and positive and negative predictive values (NPV, PPV). Analytical sensitivity is the probability that the test detects the specific mutation or those mutations that the test was intended to detect;

From the Public Health Genetics Unit, Cambridge, UK.

Mark Kroese, Public Health Genetics Unit, Strangeways Research Laboratory, Worts Causeway, Cambridge CB1 8RN, UK.

Received: March 30, 2004.

Accepted: June 19, 2004.

DOI: 10.1097/01.GIM.0000144060.84960.36

Copyright © American College of Medical Genetics. Unauthorized reproduction of this article is prohibited.

analytical specificity is the probability that the test does not detect specific mutations or mutations that are not present. Slightly differently, clinical sensitivity is the probability of a positive test result when disease is present. Clinical specificity is the probability of a negative test result when disease is absent. PPV and NPV are calculated for both analytical and clinical validity. In analytical validity, PPV is the proportion of samples with positive test results that have the mutation of interest and NPV is the proportion of samples with negative test results that do not have the mutation of interest. In the case of clinical validity, PPV is the proportion of patients with positive test results that have the disease and NPV is the proportion of patients with negative test results that do not have disease.

All these measures may be calculated by presenting the results of the test and disease (or mutation status) in a 2×2 table as shown in Table 1.

A complication arises with predictive and predisposition testing. In this situation, the "true" categorization of disease can only be made after the passage of time. It can be debated whether the 2×2 table approach is appropriate for these forms of testing.

The characteristics of a test are critically influenced by the population on which the test is performed and on the prevalence of disease in that population. The sensitivity and specificity of a test remains constant as the prevalence of disease changes, but the positive and the negative predictive values vary with disease prevalence, especially for tests of low sensitivity and specificity. This property allows the sensitivity and specificity findings from a study performed on one population to be applied to other populations with different disease prevalence. However, this assumption holds only as long as the clinical spectrum of cases in the diseased and nondiseased groups remain the same in the two populations,[10] in other words, if there is no spectrum bias or selection bias that might differentially affect the definitions of disease and nondisease.[11,12]

The likelihood ratio (LR) is also an important measure of the clinical validity of a test. It is derived from the test sensitivity and specificity and expresses the odds that a particular test result would be expected in a patient with the target disorder. For example, the positive LR is the probability of a positive test result in the presence of the disease divided by the probability of a positive test result in the absence of the disease. Because the LR is a function of sensitivity and specificity, it will also only be stable between populations if the clinical spectrum is the same.

In order to establish the analytical and clinical validity of a genetic test, controls need to be included in the assessment. For analytical validity, negative controls are samples that are known not to have the mutation of interest; whereas for clinical validity the controls are individuals that do not have the phenotype of interest according to the disease case definition. It is only through the use of such controls that an accurate assessment of specificity can be made. If possible, the analysis of the results of the tests should be blind to the disease status of participants. This will minimize test review bias.[13] Failure to use controls will critically undermine any performance measures obtained.

THE APPLICATION OF GENETIC TESTS

Genetic tests may be performed for a variety of purposes. These include the following: (1) Diagnostic testing to confirm or rule out a known or suspected genetic disorder in a symptomatic individual; (2) Predictive testing to determine the probability of asymptomatic individuals who are suspected of having an inherited disorder developing the clinical manifestations; (3) Susceptibility (or predisposition) testing to determine the risk or probability that individuals with the genetic mutation will develop a particular disease; (4) Carrier testing to identify individuals who have a gene mutation for a disorder inherited in an autosomal recessive or X-linked recessive manner; (5) Prenatal testing to determine during pregnancy whether there is an increased risk of having a child with a genetic condition; and (6) Population screening to identify asymptomatic individuals from within a particular community or a subsection of that community who have an increased chance of having a specific genetic disorder, of carrying a specific genetic predisposition to disease, or of being a carrier of a recessive genetic variant.

Genetic tests are also heterogenous in nature and the exact characteristics of a particular genetic test to be evaluated must be tightly defined. A particular genetic condition may be caused by more than one gene, *locus heterogeneity*; or by more than one variant within the gene, *allelic heterogeneity*. These variations may be due to deletions and insertions, which are not detected by routine sequencing methods. When describing a test, the gene(s) and the mutations within in it (or them) must be specified in detail. These complexities make it essential when talking about a genetic test, or undertaking its evaluation, to be clear about the following: (1) the objective and scope

Table 1
Measures of test performance

		Disease status		
		Yes	No	
Test	Positive	a (True positives)	b (False positives)	a+b (Total test positive)
	Negative	c (False negatives)	d (True negatives)	c+d (Total test negative)
		a+c (Total with disease)	b+d (Total without disease)	a+b+c+d

Sensitivity = a/(a+c); Specificity = d/(b+d); PPV = a/(a+b); NPV = d/(c+d).

Copyright © American College of Medical Genetics. Unauthorized reproduction of this article is prohibited.

of the test; (2) the population in which the test is being performed; and (3) the purpose for which the test is being performed.

The term "genetic test" is therefore shorthand to describe a test (1) to detect a particular genetic variant (or set of variants), (2) for a particular disease, (3) in a particular population, and (4) for a particular purpose. Evaluation should not proceed until and unless all four factors are formally defined; nor should it be assumed that once the characteristics of a genetic test are evaluated for one of these reasons, that the evaluation will hold or be useful for other purposes. A test's performance will, for example, vary considerably between a clinic setting and the community, in part because of the different prevalence of disease in the two populations. The clinical sensitivity and the positive predictive value of an unspecified *BRCA* test for breast cancer will, for example, depend critically on whether the test embraces both *BRCA1* and *BRCA2* and on the proportion of coding sequences that are tested for in each gene. The set of mutations that the test is designed to detect should be defined and its limitations described.

All the measures of test performance should be presented with their 95% confidence intervals. Standard statistical software packages will be able to carry out these calculations. There is a risk that some genetic test evaluations will only be able to include data on a small number of cases and hence the confidence intervals will be wide. It should become standard practice to establish appropriate sample sizes for determining acceptable specificity and sensitivity[6,14] and include these in pilots.[4]

The performance of a test during a pilot study may also be significantly better than the experience of using the same test on a standard care basis. An analogous situation is found in medical research when the performance of research studies (efficacy) is noted to be different from that seen in practice (effectiveness). For genetic tests, this can be monitored through quality assurance programs.

CONCEPTUAL ISSUES IN THE EVALUATION OF GENETIC TESTS

Epidemiology is entirely dependent on accurate case definition. An accurate definition of the reference standard by which disease status is established is therefore crucial; without such a definition it will not be possible to assign individuals to the disease positive and disease negative categories. This requirement holds for all types of tests. The usual problem is not a failure to distinguish the difference between disease and nondisease, but of (1) deciding what criteria should be used to make the distinction and (2) operationally assigning individuals to these two categories using the chosen distinction. If we wish to study the characteristics of certain tests for the diagnosis of colon cancer, there is little that is conceptually complex about what we mean by colon cancer as distinct from entities that are not colon cancer. The difficulties lie in the clinical reference standard or standards that we might employ in order operationally to partition individuals into the colon cancer or noncolon cancer groups.

For genetic disorders, there is a further problem, not usually encountered in nongenetic conditions. There appears to be genuine lack of agreement about the exact conceptualization of a genetic disease and of its definition. The "disease" is sometimes defined phenotypically by the presence of certain symptoms and signs, and at other times genotypically by reference to the mutations that give rise to the disease. Some may suggest that the definition of a genetic disease should require both the clinical manifestations and the presence of one or other mutation, whereas others will believe that the presence of either the clinical features or the mutation suffices for a definitive diagnosis.

For example in Huntington disease, the presence of the characteristic neurological and psychiatric features will certainly elicit a response that the patient "has the disease." Some authorities will also say that an asymptomatic individual with the relevant mutation "has the disease," whereas others will not and will insist that such an individual only has a predisposition to the disease. At the other end of the spectrum, it is unlikely that any asymptomatic individual shown to have a high-risk HFE genotype, for example, a C282 years homozygote, would be labeled as having hemochromatosis. However, it is problematic how one should categorize an asymptomatic individual where the probability of developing clinical manifestations is not so clear-cut. In hereditary nonpolyposis colon cancer (HNPCC), most geneticists would argue that both those with clinical (Amsterdam) criteria and those possessing associated genes, MLH1, MSH2, and PMS2, should be regarded as "having the condition"; in other words, that HNPCC should be defined by the Boolean OR joining phenotype and genotype. But is this correct? Is HNPCC to be conceptualized as either a genotypic or phenotypic diagnosis?

A genetic disease, disorder, or phenotype can be difficult to define. We believe that, for the purpose of test evaluation, the definition must be made either by reference to clinical features or by reference to genotype, and that to suggest that a disease can be defined by reference to a combination of both clinical features and genotype is conceptually unsound. We suggest that the use of phenotype is at present probably more appropriate. All genetic diseases can be defined clinically. In some, for example, tuberous sclerosis complex (TSC), cases have been formally classified "definite," "probable," and "possible" on the basis of consensus criteria.[15]

An alternative approach may be to use the genotype to define the genetic disease. But in this case, the "test" at issue would be a different technology of mutation detection than that used by the reference test to establish genotype. It would also be possible in these circumstances to evaluate the performance of a set of clinical criteria used to define the disease of interest against the reference genotype. For example, the performance of the clinical criteria used to diagnose TSC cases involving the presence of major and/or minor disease features can be assessed with reference to the genotype status. The results would indicate how well the clinical criteria performed as

Copyright © American College of Medical Genetics. Unauthorized reproduction of this article is prohibited.

a test of disease status when the reference standard is the genotype.

ANALYTICAL VALIDITY

Tests in practice rarely perform in a technically perfect manner with 100% sensitivity and specificity; false positives and false negatives will occur as well as results that cannot be interpreted with any degree of confidence. This problem is as likely to affect genetic as nongenetic tests.

Technical failures during testing may be handled analytically in many different ways. If all of these are classified as negative then the sensitivity estimate is decreased, whereas if they are all classified as positive, the specificity is decreased. If the results are not included in the assessment at all then both these test characteristics will be inflated. However, in specific conditions, where the numbers of uninterpretable test results are known not to be dependent on the underlying disease or on whether the test result would have turned out positive or negative, the estimates of sensitivity or specificity in which these are ignored will lead to unbiased estimates.[16] In clinical practice, technical failures are resolved by repeat testing or sampling. However, when formally evaluating the characteristics of a test, the repeat testing results should not be included in the calculations of test performance.

Another difficulty found with genetic testing is that mutations will be identified that cannot be confirmed definitively to be either normal variants or definitely pathological, despite the use of reference databases and the further testing of relatives. One possible method of dealing with these uncertain mutations is to treat such results as uninterpretable.

The reference standard when estimating analytical validity is another test, which has established the presence or absence of the mutation(s) of interest in the samples for evaluation. The analytical validity only describes the test's performance at identifying the stated mutations. It should be possible to achieve analytical sensitivity and specificity close to 100%.[17] The description of the analytical performance of a test should therefore include not only the testing methodologies but also the exact description of the mutations that were tested for. A change in the number and types of mutations used in an evaluation could affect the analytical results for a test.

CLINICAL VALIDITY

The concept of clinical validity is valid only if the disease definition includes the phenotype for a particular condition.

A test is usually evaluated by comparing its results on individuals known to have the disease against those without. Inevitably there will be individuals without disease who have positive test results, the false positives (Table 1); as well as those with disease whose test results are negative, the false negatives (Table 1).

The problem in the case of highly penetrant single gene disorders is that the geneticist finds it difficult to accept that those in the false-positive group (with a positive test but without disease) are "real" false positives. The test result will be interpreted as suggesting that the individual concerned has a predisposition to the disease, or that he or she "has the disease" but has failed as yet to show any manifestations. The concept of "penetrance" will be used to justify such an interpretation, and the "fault" will be laid not on the test, but on the premise that a "disease positive" individual had been wrongly assigned to a "disease negative" group, a misclassification error. The idea that an individual, without manifestations of the disease but showing a positive genetic test might in fact be a "true" or "real" false positive will be dismissed. This provides further evidence for the need to include a sufficient number of controls when evaluating the performance of a test. Controls serve to confirm whether there are any true false positives. Three reasons exist for finding a false positive: (1) the test is not one of perfect analytical specificity; (2) the correct genotype has been identified but the phenotype is not present because of reduced penetrance; (3) the disease has been clinically misclassified.

False-negative tests are more easily understood. The assumption is made that the presence of the clinical condition necessarily entails the presence of a pathogenic mutation that a test of perfect analytical validity should identify. Four reasons exist for not finding such a mutation: (1) the test is not one of perfect analytical sensitivity; (2) genes other than the one tested for are responsible for the disease (genetic heterogeneity); (3) variants other than those tested for are responsible for the disease (allelic heterogeneity); (4) the disease has been clinically misclassified.

An important contributor to the clinical misclassification of affected individuals is expressivity. Expressivity is the degree to which an inherited characteristic is expressed in a person. This is different from penetrance, which is the probability that the disease will be expressed. If fully penetrant, all individuals will show some manifestations of the disease but because of variable expressivity the severity of the manifestations will differ from patient to patient. Some of the cases will have such mild signs or symptoms that although affected they will be clinically misclassified as unaffected individuals.

False-negative results in individuals with disease may also be found when there is mosaicism, where only a proportion of cells contain a mutation. This occurs in genetic conditions caused by sporadic mutations, for example, tuberous sclerosis complex. In this case, false-negative results may occur because of the small number of cells containing the mutation in the sample being tested.[18]

If the prevalence of a genetic condition is low in the population being tested, for example, 1 in 25,000, and a relatively small number of tests have been performed (perhaps 2,000), then it is unlikely that the test will have identified a true or false positive. This will result in uninterpretable results for the performance of such a test. However, a Bayesian approach, which involves updating the prior probability distribution for the condition to a revised posterior distribution incorporating the results of a test in practice, may allow some reasonable estimate of test characteristics to be used. A detailed explanation of this approach has been published elsewhere.[19]

Copyright © American College of Medical Genetics. Unauthorized reproduction of this article is prohibited.

For many genetic diseases, knowledge of the spectrum of mutations is not complete and a particular molecular genetic test may not be able to detect all the different mutations that may be responsible for causing the disease. The idea that there might be a molecular test reference standard that could include each and every pathogenic mutation is only a theoretical possibility. A particular test will only be able to detect a proportion of the mutations present in the diseased group even if performing perfectly. These considerations have led to the development of the term "detectable" mutations.[6] Two separate concepts of clinical validity are therefore used in the literature. One by assessing the test's performance against all possible pathogenic mutations, including those not expected to be identified using that method; the other against only those mutations ("detectable mutations") that are deemed detectable using that technique based on estimates from mutation surveys and present knowledge of pathological mutations. The latter will produce higher values of clinical sensitivity and specificity.

IMPERFECT REFERENCE STANDARD

The use of a clinical reference standard will not result in a perfect evaluation even if there is consensus on the definition of disease among clinicians. This is because there will inevitably be difficulties in specifying the exact diagnostic criteria corresponding to the definition, and because clinical judgment involving intra- and interobserver variation will cause individuals to be misclassified. The degree of imperfection is nearly always unknown. Use of an imperfect reference standard to determine disease status will significantly affect test characteristics and will, provided the test and the reference standard are independent of each other, lead to an underestimate of its sensitivity and specificity. The test's sensitivity will be most accurately estimated when the disease prevalence is high, its specificity when disease prevalence is low.[20] The positive and negative predictive values, however, may be biased either positively or negatively.[21] Published methods are available to establish unbiased estimates of sensitivity and specificity when the accuracy of the reference standard is not known but these can be complicated and difficult to use.[22]

Another approach that has been used to address the issue of an imperfect reference standard is discrepant analysis.[23,24] This type of analysis involves looking at the group of false positives (Cell b, Table 1). If investigators believe that the test under evaluation is more sensitive for predicting the presence or risk of disease than the reference standard, it is possible to perform a second different test on the false positives. A proportion of Cell b cases will be positive with this second test and the analysis is adapted by defining a new reference standard to allow these to be added to Cell a (Table 1) cases. This form of discrepant analysis results in estimates of the sensitivity and specificity that are greater than their "true" value and invariably favors the new diagnostic test under assessment. An example that illustrates this method involves the comparison of a plasmid based ligase chain reaction (LCR) test to the reference standard of cell culture for *C trachomatis*. The false positives for the LCR test underwent further testing with a PCR test, which identified a proportion of these false positives to be true positives. The 2×2 table was then modified to include these true positives and the resulting sensitivity and specificity values were higher.[25] We, in common with others, believe this to be a biased approach and should not be used in the assessment of genetic tests. If a further test is to be used in the assessment process then it should be applied to all cases, not just the false positives. The performance of the combined tests can then be assessed appropriately. Discrepant analysis fails to adhere to the important principle, that the reference standard should not include tests that are dependent on the new test being evaluated.[25–27]

Other approaches used for the problem of imperfect reference standards include the following: (1) Composite reference standards: the results of several, imperfect tests are combined to define a composite reference standard (which must exclude the new test).[28] This approach uses several sources of information sequentially, so avoiding redundant testing. It does not depend on the new test's results, in contrast to discrepant analysis. (2) Latent class analysis: this method recognizes that the true status of an individual is unknown (or latent) and relates the observed diagnostic test results to the unknown truth using a statistical model.[29] A minimum of three imperfect reference standards must be performed on each specimen. The basic method assumes that each test is independent of the others. This assumption is not always true but there are ways of dealing with conditional dependence statistically.

CONCLUSION

We have discussed some of the issues involved in the evaluation of genetic tests, concentrating on two particular test performance measures, analytical and clinical validity. However, this is only part of the much larger evaluation process and many other factors, such as clinical utility, need to be considered as well.[6] This process can be complex but with appropriate support and resources, should after several cycles become quicker and more efficient. The subject is as yet poorly developed, but as experience is gained in the regular evaluation of genetic tests new insights will emerge and greater understanding will result. The evaluation of predictive and predisposition genetic tests also requires further development.

The limitations of our genetic knowledge and technical abilities means that for the moment there are likely to be gaps in the information needed to complete a thorough evaluation of many genetic tests. In particular the inability to identify all disease-related mutations makes it difficult to estimate clinical validity. The tension between wanting to use new technology in clinical practice and the delay involved in evaluating a test needs to be acknowledged. This is especially true with the fast pace of technical development in genetic research and the small number of samples for testing in many genetic conditions. However, this should not be accepted as a reason for bypassing the evaluation process. Evaluation may be imperfect and results incomplete, but failure to perform adequate assess-

Copyright © American College of Medical Genetics. Unauthorized reproduction of this article is prohibited.

ments of new tests will reduce the quality of health care and have a detrimental effect on the public health.

ACKNOWLEDGMENTS

We are grateful to Professor Wylie Burke for her comments on drafts of this review.

References

1. Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. Getting better but still not good. *JAMA* 1995;274:645–651.
2. Bogardus ST, Jr., Concato J, Feinstein AR. Clinical epidemiological quality in molecular genetic research: the need for methodological standards. *JAMA* 1999;281: 1919–1926.
3. Nierenberg AA, Feinstein AR. How to evaluate a diagnostic marker test. Lessons from the rise and fall of dexamethasone suppression test. *JAMA* 1988;259:1699–1702.
4. Final Report of the Task Force on Genetic Testing in the United States. Promoting Safe and Effective Genetic Testing in the United States. Holtzman NA and Watson MS. 1–82. 1997. The National Human Genome Research Institute.
5. Zimmern R. What is genetic information: Whose hands on your genes? *Genet Law Mon* 2001;1:9–13.
6. ACCE. Population-Based Pre-natal Screening for Cystic Fibrosis via Carrier Testing. Haddow JE and Palomaki G. 2002. Office of Genomics and Disease Prevention, CDC.
7. Secretary's Advisory Committee on Genetic Testing. Enhancing the Oversight of Genetic Tests, Recommendations of the SACGT. 1–32. Bethesda, Md: National Institutes of Health; 2000.
8. Secretary's Advisory Committee on Genetic Testing. Development of a Classification Methodology for Genetic Tests: Conclusions and Recommendations of the Secretary's Advisory Committee on Genetic Testing. 1–7. Bethesda, MD: National Institutes of Health; 2001.
9. Burke W, Atkins D, Gwinn M, Guttmacher A, Haddow J, Lau J et al. Genetic test evaluation: information needs of clinicians, policy makers, and the public. *Am J Epidemiol* 2002;156:311–318.
10. Brenner H, Gefeller O. Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. *Stat Med* 1997;16:981–991.
11. Knottnerus JA, Van Weel C. General introduction: evaluation of diagnostic procedures. In Knottnerus JA, ed. The Evidence Base of Clinical Diagnosis. London: BMJ Books; 2002:1–18.
12. Mulherin SA, Miller WC. Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation. *Ann Intern Med* 2002;137:598–602.
13. Begg CB. Biases in the assessment of diagnostic tests. *Stat Med* 1987;6:411–423.
14. Alonzo TA, Pepe MS, Moskowitz CS. Sample size calculations for comparative studies of medical tests for detecting presence of disease. *Stat Med* 2002;21:835–852.
15. Roach ES, Gomez MR, Northrup H. Tuberous sclerosis complex consensus conference: revised clinical diagnostic criteria. *J Child Neurol* 1998;13:624–648.
16. Begg CB, Greenes RA, Iglewicz B. The influence of uninterpretability on the assessment of diagnostic tests. *J Chronic Dis* 1986;39:575–584.
17. 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;5:440–443.
18. Kwiatkowska J, Wigowska-Sowinska J, Napierala D, Slomski R, Kwiatkowski DJ. Mosaicism in tuberous sclerosis as a potential cause of the failure of molecular diagnosis. *N Engl J Med* 1999;340:703–707.
19. Smith JE, Winkler RL, Fryback DG. The first positive: computing positive predictive value at the extremes. *Ann Intern Med* 2000;132:804–809.
20. Boyko EJ, Alderman BW, Baron AE. Reference test errors bias the evaluation of diagnostic tests for ischemic heart disease. *J Gen Intern Med* 1988;3:476–481.
21. Valenstein PN. Evaluating diagnostic tests with imperfect standards. *Am J Clin Pathol* 1990;93:252–258.
22. Walter SD,. Irwig LM. Estimation of test error rates, disease prevalence and relative risk from misclassified data: a review. *J Clin Epidemiol* 1988;41:923–937.
23. Green TA, Black CM, Johnson RE. Evaluation of bias in diagnostic-test sensitivity and specificity estimates computed by discrepant analysis. *J Clin Microbiol* 1998;36: 375–381.
24. Lipman HB,. Astles JR. Quantifying the bias associated with use of discrepant analysis. *Clin Chem* 1998;44:108–115.
25. Hadgu A. The discrepancy in discrepant analysis. *Lancet* 1996;348:592–593.
26. Hadgu A. Discrepant analysis: a biased and an unscientific method for estimating test sensitivity and specificity. *J Clin Epidemiol* 1999;52:1231–1237.
27. McAdam AJ. Discrepant analysis: how can we test a test? *J Clin Microbiol* 2000;38: 2027–2029.
28. Alonzo TA,. Pepe MS. Using a combination of reference tests to assess the accuracy of a new diagnostic test. *Stat Med* 1999;18:2987–3003.
29. Formann AK, Kohlmann T. Latent class analysis in medical research. *Stat Methods Med Res* 1996;5:179–211.

Copyright © American College of Medical Genetics. Unauthorized reproduction of this article is prohibited.