How can the evaluation of genetic tests be enhanced? Lessons learned from the ACCE framework and evaluating genetic tests in the United Kingdom

Simon Sanderson MRCP, MFPHM¹, Ron Zimmern FRCP FFPHM², Mark Kroese MRCGP, MFPHM², Julian Higgins PhD², Christine Patch PhD³, Jon Emery MRCGP, PhD⁴

Advances in genetic technology are increasing the availability of genetic tests, not only for rare single gene disorders, but also for common diseases such as breast and colo-rectal cancer. Before there can be widespread uptake of these tests, they must be evaluated to confirm the benefits of their use. But how should genetic tests be evaluated, given the speed at which new tests are emerging? One highly influential approach is the analytic validity, clinical validity, clinical utility and ethical, legal and social issues (ACCE) framework, which has provided a benchmark for the evaluation of genetic tests. The approach has been adopted and adapted by the United Kingdom Genetic Testing Network, with the help of the Public Health Genetics Unit in Cambridge, to evaluate new genetic tests for use in the National Health Service. We discuss a number of conceptual, methodological, and practical issues concerning the evaluation of genetic tests, based on lessons learned from applying the ACCE framework and from the UK experience, and make a number of recommendations to further strengthen the evaluation of genetic tests. **Genet Med 2005;7(7):495–500.**

Key Words: Genetic tests, evaluation, diagnosis, prediction

Advances in genetic technology are increasing the availability of genetic tests, not only for rare single gene disorders, but also for common diseases such as breast and colo-rectal cancer. Before there can be widespread uptake of these tests, they must be evaluated to confirm the benefits of their use. But how should genetic tests be evaluated?

One approach has been the ACCE model framework for evaluating emerging genetic tests.¹ The model takes its name from the four components evaluated: analytic validity, clinical validity, clinical utility and ethical, legal and social issues. The framework has become a benchmark for evaluating genetic tests. Five disorders were chosen to pilot the system and reports are now available for them all.² The project was completed in 2004. The Office of Genomics and Disease Prevention (OGDP) has recently launched a new three-year project, the *Evaluation of Genomic Applications*

in Prevention and Practice (EGAPP), following on from using the ACCE framework.³

The United Kingdom Context

The United Kingdom Genetic Testing Network Steering Group, which informs the commissioning and provision of genetic tests in the NHS, has endorsed and adapted the ACCE core principles to produce a “Gene Dossier” for evaluating genetic tests in the National Health Service.⁴ The Public Health Genetics Unit in Cambridge has had a leading role in the development and application of the Gene Dossier. The Gene Dossier was introduced in 2003, and since its introduction, approximately 30 genetic tests have been formally evaluated by a working group of the Genetic Testing Network. One of the authors (MK) is a member of the evaluation group. The Dossier’s emphasis has been on emerging genetic tests for rare inherited disorders. The Public Health Genetics Unit has also used the ACCE framework to complete a review of the role of genetic testing in familial hypercholesterolaemia (FH). This has given us the opportunity to use the ACCE framework in practice.

This discussion paper explores how the evaluation of genetic tests could be enhanced, based on lessons learned from applying the ACCE framework and our own experience of evaluating genetic tests in the UK.

From the ¹University of Cambridge and Cambridge Genetics Knowledge Park, Cambridge UK, ²Public Health Genetics Unit, Cambridge Genetics Knowledge Park, Cambridge UK, ³Public Health Genetics Unit, Cambridge Genetics Knowledge Park and MRC Biostatistics Unit, Cambridge UK, ⁴Public Health Sciences and Medical Statistics, School of Medicine, University of Southampton, Southampton UK, ⁵Discipline of General Practice, University of Western Australia, Perth Australia.

Dr Simon Sanderson, MRCP, MFPHM; Strangeways Research Labs, Wort’s Causeway, Cambridge CB1 8RN, UK.

Received: January 10, 2005.

Accepted: May 25, 2005.

DOI: 10.1097/01.gim.0000179941.44494.73

THE ACCE METHODOLOGY

We believe that the four ACCE categories for evaluation are the correct ones, following a logical sequence from the laboratory to the clinical setting. However, generic methodological guidance for addressing each category is currently not available, although each of the published ACCE reviews and the ACCE website provide examples of different approaches². Methodological guidance will be invaluable if ACCE is to develop into a transferable system for evaluating genetic tests.

Obtaining and appraising evidence

Evaluating diagnostic tests is methodologically challenging and the quality of many published evaluations is poor.^{5–8} Although some questions are more amenable to standard evidence-review techniques than others, systematic review of the literature should be the standard method of obtaining evidence, supplemented with meta-analysis when applicable (for example, to combine indices of diagnostic accuracy, such as sensitivity). The type of evidence used to reach specific conclusions should be set out, along with an assessment of the quality of reporting, studies' susceptibility to bias and where possible, placed within an hierarchy of evidence.^{9–13} Methods should build on guidance already available for the conduct of systematic reviews and meta-analyses of diagnostic studies and the evaluation of population genetic screening for cancer.^{9,11,14,15} A major problem for many new and emerging genetic tests is that the evidence base is limited in scope, making it difficult to complete a full ACCE evaluation. This lack of data was one of the main lessons learned from the Gene Dossier process in the UK. Nevertheless, the evaluation process should help to identify gaps in the evidence base and help set the agenda for further research.

Small numbers, sample size and power

For new tests of rare disorders, the very small numbers involved (e.g., low prevalence, few total number of tests performed, and a lack of false positive or false negative results) may make it difficult to calculate specificity and predictive values. Confidence intervals for these indices are also usually wide. The “zero numerator” problem for false positive or false negative results can be dealt with in a number of ways, including the “rule of three” approach and Bayesian methods¹⁶. Supporting guidance for their application within evaluative reviews should be developed. Again, meta-analysis may help to increase power and provide pooled estimates of test performance. Prospective meta-analysis may be a useful tool for incorporating the results of new studies as they are published.

DISORDER AND SETTING

The term “genetic test” is shorthand to describe a test to detect (1) a particular genetic variant (or set of variants), (2) for a particular disease, (3) in a particular population and (4) for a particular purpose¹⁷. The ACCE review recognizes the importance of these four characteristics of genetic tests by plac-

ing pertinent questions at the beginning. More information about the scope and purpose of a test would strengthen this section further, because these factors influence the estimation and interpretation of validity and utility. Determining the genetic variants known to be associated with a disease entity, what we term characterizing the genotype of interest, is a key step. This genotype could be defined in three ways:

- All known and unknown variants
- All known variants
- Specific selected variants

For example, there are more than 900 known mutations of the *CTFR* gene associated with cystic fibrosis.¹⁸ A genetic test for cystic fibrosis could aim to identify all of these known variants, or only a smaller subset of specific, selected mutations (often referred to as a panel). The decision to use specific variants is often based on the frequency of those variants in certain populations; the United States panel for cystic fibrosis currently uses a panel of 25 mutations.¹⁹

Subsequent steps are defining who is to be tested (populations, families or individuals) and defining the purpose(s) of testing (diagnostic, screening, predictive, susceptibility, or carrier testing). The purpose of the test is critical to any evaluation and any multi-purpose genetic test should be thoroughly evaluated with respect to each of its core functions.

Although some of these issues are dealt with in subsequent sections of the ACCE review, we believe that placing them together at the beginning makes more sense. Because of the importance of these issues, we recommend that the disorder and setting section should become the first domain of the formal ACCE framework.

ANALYTIC VALIDITY

Analytic validity defines the ability of a test to accurately and reliably measure the genotype of interest: the genetic variant(s) that the test is aiming to identify.¹ This part of the evaluation is concerned with assessing test performance in the laboratory as opposed to the clinic. Explicit specification of the genotype of interest is needed because the estimation of analytic validity is both method- and mutation-specific. This is an important factor when comparing or combining results from different laboratories.

One important problem for analytic validity concerns those tests where the genotype of interest cannot be easily specified, usually when there is extensive allelic and/or locus heterogeneity or when mutation-scanning methods are used. One commonly used mutation-scanning method, Single Strand Conformation Polymorphism analysis (SSCP), can detect the presence of a mutation in a sample but cannot determine its position or specify its type.²⁰ Subsequent sequencing is often required to further characterize the features of the detected variant and to determine whether it is a known variant of clinical significance or not. This means that it may be difficult to quantify analytic validity unless the definition currently used

in the ACCE framework is adopted or an alternative can be identified.

The choice of an appropriate reference standard is also essential for determining validity.^{9,21–23} Additional questions about reference standards should be sought by the review.

Finally, the concepts of analytic validity and quality assurance are appropriately linked within the ACCE framework. Quality assurance aims to ensure that test results are reliable and reproducible and usually include internal and external control assessments within a quality management framework. For many genetic tests, especially low-volume tests for rare conditions, quality assurance will be the main way of assessing analytic validity. However, many emerging tests may not have a comprehensive external quality assurance system. The ACCE process can therefore help to identify gaps, which would be of interest to clinicians, policy-makers and regulators and to facilitate development of relevant quality assurance programs.²⁴

CLINICAL VALIDITY

This is defined as the ability of the test to detect or predict the phenotype of interest.¹ The definition of the phenotype of interest and its determination (for example, using clinical diagnostic criteria or a biochemical marker) is a key task. The goal is to accurately determine those individuals who have (or will develop) the phenotype of interest. It is critical that the definition of the phenotype should not also include the results of genetic testing. For the purposes of test evaluation, disorders should be defined with reference to the phenotype or the genotype but not both.¹⁷ Definition of a phenotype raises different issues depending on whether the purpose of the test is, for example, to diagnose an individual suspected of having a specific disease, to predict future occurrence of a disease, or to determine the carrier status of a relative. The problem of small numbers described above in the methodology section is also relevant for assessing clinical sensitivity, specificity and predictive values, especially for tests of very rare conditions. Here we discuss a number of important issues for evaluating clinical validity.

Diagnosis versus prediction

Although the ACCE definition of clinical validity includes both detection and prediction, the current questions are based on a diagnostic paradigm. There are limitations when applying this approach to risk prediction because of the need to incorporate the complexities of genetic testing and especially the dimension of time.²⁵ To detect genetic variants that confer an increased risk of developing the phenotype during a defined time period requires data from cohort studies; cross-sectional studies are generally inappropriate. Applying indices of diagnostic accuracy (such as sensitivity and specificity) are less meaningful, where the aim is estimating absolute risks for individuals. Because of these problems we believe that a separate approach would be more appropriate for evaluating predictive tests.

A number of the published ACCE reviews have addressed predictive testing but have had difficulties in applying a consistent approach.²⁶ For example, in the hemochromatosis review, clinical sensitivity and specificity have been determined using a cross-sectional design, with the genetic test performed in patients who already have primary iron overload. The review's authors state that this is not the ideal design, which would be a cohort study of persons free of disease, tested for the mutation and then followed up. Although they believe that this approach is not feasible for hemochromatosis, they conclude that a more appropriate group should be sought to validate their estimates based on cross-sectional data. In the breast and ovarian cancer review, no answers are given for questions 18, 19 and 23 (which deal with clinical sensitivity, specificity and predictive values, respectively). This may be because of the problems in obtaining data. In the colo-rectal cancer review, clinical sensitivity, specificity and predictive values are reported for mutations detected in persons who already have familial or nonfamilial colo-rectal cancer. The lifetime risk of developing colo-rectal cancer with certain genetic variants is also reported. However, this is not the same as estimating an individual's time-specific risk of developing colo-rectal cancer from a specific genetic test result because tests do not perform perfectly, especially when applied in a routine clinical setting. The best treatment of predictive testing is in the venous thrombo-embolism review, which uses data from cohort studies. A consistent methodological approach for dealing with predictive testing is therefore needed.

Study populations for evaluating clinical validity

The selection of a clinically relevant cohort is a fundamental concern for any study evaluating genetic tests.^{8,9} Wherever possible, this cohort should be as similar as possible to the patients who will be tested in practice, in terms of the health care setting where the test will be performed, the clinical spectrum of the target disorder, and the inclusion of appropriate controls.

Empirical research has shown that studies recruiting diseased patients separately from those without disease (such as using a group of healthy controls) overestimate test performance when compared with studies recruiting a cohort of patients unselected by disease status that were representative of the test's clinical population.⁸ However, it also needs to be acknowledged that detailed evaluation of tests of very rare disorders may be challenging because of the difficulties in defining a suitable study group.

"False" versus "true" test results

Designating genetic test results as either "true" or "false" creates additional problems because it is linked to concepts of penetrance and expressivity and to technical issues of testing.²⁷ False positive results can be caused by:

- Sample mix-up
- Technical errors with the test
- Imperfect analytic specificity
- Reduced penetrance

- Clinical misclassification

False negative results can be caused by:

- Sample mix-up
- Technical errors with the test
- Imperfect analytic sensitivity
- Locus and allelic heterogeneity
- Clinical misclassification

In single gene disorders known to be highly penetrant, it may be difficult to accept that false positive results are “real” false positives, resulting in interpretation problems. After excluding technical errors, does this mean that the individual concerned has a predisposition to the phenotype, or that they have the phenotype but so far have failed to show any clinical manifestations as a result of reduced penetrance? It may be hard to accept that the fault lies in the testing process and not on the premise that a disease positive individual has been wrongly assigned to a disease negative group (a misclassification error). The idea that an individual, without phenotypic manifestations but showing a positive genetic test might in fact be a “real” false positive may be dismissed. False negative tests are more easily understood. An 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; however, this assumption may be false (for example, as a result of a phenocopy). We believe that greater emphasis is needed on the evaluation of both false positive and false negative results because they can have important clinical consequences for patients and their families.

Individuals and families

For some applications of genetic testing, the unit of analysis is the family and not simply individuals. For example, in familial breast-ovarian cancer there are three key steps:

- Deciding whether the woman’s *family* should be tested, based on an assessment of the family history
- Mutation scanning in an affected family member to identify a relevant mutation segregating in that family
- Direct gene testing in apparently unaffected family members.

This three-step process is analogous to series screening, which usually results in an increase in specificity. This is clearly brought out in the section on series testing in the ACCE reviews of breast and ovarian cancer and colo-rectal cancer.²⁶ However, the impact of combining the results from these different stages is not adequately dealt with in the section on clinical validity, where no data are provided on the conditional probabilities involved. For example, in the colo-rectal cancer review, the section on predictive values present results from consecutive colo-rectal cancer patients, not accounting for family history. We believe that calculating clinical validity indices in individual, unaffected family members should take into account the probability of them being offered a direct gene test, which is dependent on the family being scanned for seg-

regating mutations, which is dependent on an assessment of the familial risk of an inherited cancer syndrome. The validity results should be reported at each of these three stages along with the combined results.

A related problem here is that mutation-scanning tests appear not to perform as well as direct gene tests. This is because most conditions demonstrate extensive allelic heterogeneity, so a search for a mutation anywhere within or near the relevant gene (or locus) is required and not all variants are known or detectable. Current gene scanning methods (such as SSCP) are laborious, expensive and often cannot differentiate between pathogenic and nonpathogenic variants.²⁰ Subsequent direct sequencing is required to determine the location of the mutation. Direct gene tests often appear to perform better because the prior probability of detection is greatly increased.

Determination of clinical validity in different population groups

The clinical validity of a test can be directly related to the type and number of mutations included and the populations tested, especially where there ethnic differences.¹⁹

This US pan-ethnic panel for cystic fibrosis is one example where a single panel of 25 relevant mutations has been chosen for the entire population, rather than creating separate panels for specific ethnic groups or modifying the testing process to incorporate further testing and assessment in specific subgroups. This panel identifies 77% of affected Caucasians and 49% of affected Asian Americans.²⁸

CLINICAL UTILITY

This is defined as the likelihood that the test will lead to an improved outcome, and incorporates assessment of the risk and benefits of genetic testing, as well as economic evaluation.¹ This is perhaps the most important aspect of the evaluation: will the results of genetic testing alter clinical management? Will those tested gain net benefit? This is important because the numerical indices of analytic and clinical validity do not provide all the necessary information for assessing clinical utility.

Tests that perform superbly in the laboratory may not necessarily be useful in clinical practice and tests that perform relatively poorly in the laboratory may prove to be very useful clinically, because of the clinical impact. It is also important to recognize that tests used in routine clinical practice often do not perform as well as in research programs (the gap between efficacy and effectiveness). In clinical practice, there are also trade-offs to be made between sensitivity and specificity, especially when tests are used for different purposes.

Some of these issues can be illustrated using genetic testing in FH, an autosomal dominant condition with a population prevalence of around 1 in 500 and is fully penetrant by adolescence. It substantially increases the risk of coronary heart disease and stroke.^{29,30} The majority of cases are caused by mutations in the low-density lipoprotein receptor gene; there are over 700 known mutations worldwide.

Current UK diagnostic criteria for FH are based on clinical features, family history, and the results of lipid testing. Individ-

uals are subsequently categorized as definite FH or possible FH.²⁹ The advantage of a genetic test is that it can establish definitively whether a person has FH or not by detecting the presence of an FH-causing mutation. However, the clinical sensitivity of genetic testing is variable (35-80%; reference standard clinical phenotype), and the positive predictive values are poor, varying with the pretest probability in those tested.³¹ So, a person with *definite* FH apparently gains little from genetic testing because the diagnosis has already been made; their clinical management will not be altered by the result. However, this ostensibly diagnostic test result in an individual also has implications for first-degree relatives because of the inheritance pattern.

For a person with *possible* FH, a positive test result may be more useful because the diagnosis could be determined definitively, even though the reported sensitivity of testing in this group is as low as 20%; again, there are also implications for relatives. Thus a final policy decision on using genetic tests must depend on judgment and values that go beyond the quantitative assessment of test performance. This example also highlights why separate evaluations are required when evaluating genetic tests for different purposes.

Establishing the clinical context

The assessment of clinical utility should be strengthened in the ACCE framework, particularly in recognizing that genetic testing is a complex process and is one component of an overall complex intervention. This would be greatly helped by a clear description of where genetic testing fits into clinical management, using a clinical guideline or algorithm, setting out referral pathways, assessment criteria, detailing who is requesting the genetic test and why. This establishes the relationships between clinical diagnosis, nongenetic diagnostics and genetic testing, and emphasizes the inherent complexity of genetic testing. This kind of analysis should include assessment of genetic testing instead of nongenetic testing or in addition to nongenetic testing, with estimation of relevant benefits, harms and costs.

ETHICAL, LEGAL AND SOCIAL ISSUES (ELSi)

These issues are perhaps the most difficult section to deal with in the context of evaluating genetic tests. This is reflected in our own experience with FH and in the available ACCE reviews (e.g., the chapter on ELSi in the colorectal review is very brief, concentrating on the impact of test results in individual patients).²⁶ One reason for the difficulty is the heterogeneity of genetic tests: ELSi for a population screening test are different from those for a diagnostic test for a rare disease. It is very difficult to assess ELSi for specific genetic tests, especially when those tests may be very new or for very rare disorders.

While there is an extensive literature about ELSi and genetic testing, much of the discussion is related to population screening or to general principles (e.g., counseling, patient information needs, informed consent, impact on insurance), rather than specific issues for individual tests.^{14 14,24,32-39} It is difficult

to know whether, and how, to apply evidence from these general sources to individual genetic tests. As a result, ELSi may not be adequately assessed because of these difficulties. There is therefore a need for further work to define the most useful and relevant questions for emerging genetic tests in an evaluation context. The work of Burke and colleagues and the UK Department of Health's Advisory Committee on Genetic Testing *Report on Genetic Testing for Late Onset Disorders* may provide additional help in developing more focused questions.^{40,41}

PRIORITIZING TESTS FOR ACCE-TYPE REVIEWS

Given the speed at which new genetic tests are emerging, the current ACCE framework could not be easily applied to all of these tests because it would be too slow, expensive and impractical. Thus, tests may be adopted into clinical practice before they have been properly evaluated. While this is not a new phenomenon, it highlights the need for ways of prioritizing tests to review and for different levels of evaluation. The UK Genetic Testing Network has found that the information needs for evaluating genetic tests to be used in large populations versus those in very small, well-defined groups are very different. This raises the question of what the balance should be between the degree of detail required to assess test performance, the available resources, and the negative impact of not providing the test in the absence of evidence.

For a population-based screening test, the current comprehensive ACCE review process is entirely appropriate but for low volume tests, a more streamlined (but rigorous) evaluation should be more feasible. Within particular health care economies, thresholds to trigger different levels of evaluation could be developed.

There is also the question of whether one framework can adequately cover all of the relevant issues for all types of genetic test; a one size fits all approach may be too ambitious to deal with all the complexities of genetic testing, as we have argued for predictive testing. We believe that it may be beneficial to develop a series of evaluation frameworks for tests with different purposes, as well as approaches for streamlined assessments (such as the UK's Gene Dossier). Greater international collaboration, building on the Human Genome Epidemiology Network (HuGENet™) and models such as The Cochrane Collaboration, would allow workload to be shared and capacity to be increased. Individual countries would then only need to adapt results to their local health economies and regulatory structures.

CONCLUSIONS

The ACCE model is an important development in evaluating genetic tests and has much to commend it. The four domains for evaluation follow a logical and comprehensive route from the laboratory to the patient, population, and society; these basic concepts should remain unaltered. We recommend that the section on disorder and setting should become the first part of the formal ACCE evaluation process. We have de-

scribed a number of difficulties in applying the framework and suggested ways in which the evaluation of genetic testing might develop in the future. These include: better definition of the molecular, clinical and public health contexts; agreeing methodologies required to answer review questions; using different approaches to cope with the heterogeneity of genetic tests; strengthening the assessment of clinical utility; deciding when to apply comprehensive or streamlined review processes; and developing international collaboration. These changes should help evaluation to keep up with the rapidity of emerging tests and allow tests with major implications to be thoroughly evaluated when necessary.

References

- Haddow J, Palomaki G. ACCE: A Model Process for Evaluating Data on Emerging Genetic Tests. In *Human Genome Epidemiology* Khoury M, Little J, Burke W, eds. Oxford: Oxford University Press 2004;217–233.
- Office of Genomics and Disease Prevention. *A Model System for Collecting, Analyzing and Disseminating Information on Genetic Tests*. <http://www.cdc.gov/genomics/activities/fbr.htm> Accessed December 18, 2004.
- Office of Genomics and Disease Prevention. *Evaluation of Genomic Applications in Practice and Prevention*. <http://www.cdc.gov/genomics/gtesting/egapp.htm>. Accessed August 18, 2005.
- Department of Health. *United Kingdom Genetic Testing Network*. www.ukgtn.org. Accessed December 18, 2004.
- Feinstein AR. Misguided efforts and future challenges for research on “diagnostic tests”. *J Epidemiol Community Health* 2002;56:330–332.
- 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.
- Knottnerus JA, van Weel C, Muris JW. Evaluation of diagnostic procedures. *BMJ* 2002;324:477–480.
- Gluud C, Gluud LL. Evidence based diagnostics. *BMJ* 2005;330:724–726.
- Deeks JJ. Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests. *BMJ* 2001;323:157–162.
- Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Clin. Chem. Lab Med* 2003;41:68–73.
- Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann. Intern. Med* 2003;138:W1–12.
- Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 2000;283:2008–2012.
- Evans D. Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. *J. Clin. Nurs* 2003;12:77–84.
- Goel V. Appraising organised screening programmes for testing for genetic susceptibility to cancer. *BMJ* 2001;322:1174–1178.
- Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Clin. Chem. Lab Med* 2003;41:68–73.
- Smith JE, Winkler RL, Fryback DG. The first positive: computing positive predictive value at the extremes. *Ann. Intern. Med* 2000;132:804–809.
- Kroese M, Zimmern RL, Sanderson SP. Genetic Tests and their Evaluation: Can we answer the key questions? *Genet. Med* 2004;6:475–480.
- Cystic Fibrosis Genetic Analysis Consortium. *Cystic Fibrosis Mutation Database*. <http://www.genet.sickkids.on.ca/cftr/>. Accessed August 18, 2005.
- Palomaki GE, FitzSimmons SC, Haddow JE. Clinical sensitivity of prenatal screening for cystic fibrosis via CFTR carrier testing in a United States panethnic population. *Genet. Med* 2004;6:405–414.
- Humphries SE, Gudnason V, Whittall R, Day IN. Single-strand conformation polymorphism analysis with high throughput modifications, and its use in mutation detection in familial hypercholesterolemia. International Federation of Clinical Chemistry Scientific Division: Committee on Molecular Biology Techniques. *Clin. Chem* 1997;43:427–435.
- Valenstein PN. Evaluating diagnostic tests with imperfect standards. *Am. J. Clin. Pathol* 1990;93:252–258.
- 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.
- 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.
- 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.
- Evans JP, Skrzynia C, Burke W. The complexities of predictive genetic testing. *BMJ* 2001;322:1052–1056.
- Office of Genomics and Disease Prevention. *ACCE Model System; draft genetic test reviews*. <http://www.cdc.gov/genomics/gtesting/ACCE/fbr.htm>. Accessed August 18, 2005.
- Jimenez C, Gagel RF. Genetic testing in endocrinology: lessons learned from experience with multiple endocrine neoplasia type 2 (MEN2). *Growth Horm. IGF. Res.* 2004;14 Suppl A:S150–157.
- Watson MS, Cutting GR, Desnick RJ, Driscoll DA, Klinger K, Mennuti M et al. Cystic fibrosis population carrier screening: 2004 revision of American College of Medical Genetics mutation panel. *Genet. Med* 2004;6:387–391.
- Marks D, Thorogood M, Neil HA, Humphries SE. A review on the diagnosis, natural history, and treatment of familial hypercholesterolaemia. *Atherosclerosis* 2003;168:1–14.
- Mortality in treated heterozygous familial hypercholesterolaemia: implications for clinical management. Scientific Steering Committee on behalf of the Simon Broome Register Group. *Atherosclerosis* 1999;142:105–112.
- Humphries SE, Galton D, Nicholls P. Genetic testing for familial hypercholesterolaemia: practical and ethical issues. *QJM* 1997;90:169–181.
- Burke W, Zimmern RL. Ensuring the appropriate use of genetic tests. *Nat. Rev. Genet* 2004;5:955–959.
- Burke W, Coughlin SS, Lee NC, Weed DL, Khoury MJ. Application of population screening principles to genetic screening for adult-onset conditions. *Genet. Test* 2001;5:201–211.
- Godard B, Raeburn S, Pembrey M, Bobrow M, Farndon P, Ayme S. Genetic information and testing in insurance and employment: technical, social and ethical issues. *Eur. J. Hum. Genet.* 2003;11 Suppl2:S123–142.
- Godard B, ten Kate L, Evers-Kiebooms G, Ayme S. Population genetic screening programmes: principles, techniques, practices, and policies. *Eur. J. Hum. Genet.* 2003;11 Suppl 2:S49–87.
- Dickenson DL. Can children and young people consent to be tested for adult onset genetic disorders? *BMJ* 1999;318:1063–1065.
- Nuffield Council on Bioethics. *Genetic Screening: Ethical Issues*. 1993.
- Parker M, Lucassen A. Working towards ethical management of genetic testing. *Lancet* 2002;360:1685–1688.
- Zimmern RL. Genetic testing: a conceptual exploration. *J. Med. Ethics* 1999;25:151–156.
- Burke W, Pinsky LE, Press NA. Categorizing genetic tests to identify their ethical, legal, and social implications. *Am. J. Med. Genet* 2001;106:233–240.
- Advisory Committee on Genetic Testing Health Departments of the United Kingdom. *Report on Genetic Testing for Late Onset Disorders*. 1998.