phg foundation
making science work for health

MOVING BEYOND ACCE: An Expanded Framework for Genetic Test Evaluation

A paper for the United Kingdom Genetic Testing Network

Wylie Burke and Ron Zimmern

September 2007

This report is the result of work funded by the Department of Health for the UK Genetic Testing Network (UKGTN).

Dr Wylie Burke is Professor and Chair of the Department of Medical History and Ethics at the University of Washington, and Principal Investigator of the University of Washington Center for Genomics and Healthcare Equality.

Department of Medical History & Ethics
A204 Health Sciences Center
University of Washington
Box 357120
Seattle
WA 98195-7120
USA

www.washington.edu/mhedept

Dr Ron Zimmern is Executive Director of the PHG Foundation. an independent charity that works with partners to achieve better health through the responsible and evidence-based application of biomedical science

PHG Foundation
Strangeways Research Laboratory
Worts Causeway
Cambridge
CB1 8RN
UK

www.phgfoundation.org

The cover image is taken from the US Centers for Disease Control (CDC) Public Health Image Library at www.phil.cdc.gov; photographer Jim Gathany (2004).

Contents

1. Introduction

1.1 The purpose of this paper is to propose an approach to the evaluation of genetic tests that expands on and moves beyond the ACCE framework. It clarifies certain concepts key to the evaluation process, and proposes the use of measures of health quality in the evaluation of a genetic test and associated services.

1.2 The key concepts are set out in brief below. They are:

(a) the need to define separately an assay and a test and to distinguish between them

(b) the need to separate two distinct properties of clinical validity, gene-disease association and clinical test performance.

(c) the need to define test purpose as the initial step in genetic test evaluation

(d) the relevance to genetic test evaluation of genotype penetrance and geno types as necessary or non-necessary causes of disease

(e) dimensions of quality and their application to genetic test evaluation

1.3 These concepts build on the ACCE framework [1]; the definition of a genetic test and the distinction between test and assay proposed by Zimmern and Kroese [2]; the formal definition in the audit and quality literature of the effectiveness of an intervention as the extent to which it meets the objective (purpose) for which it was designed, and of the quality of an intervention as the extent to which it meets the standards that were set for it [3, 4]; Donabedian's framework for the dimensions of health care quality [3]; the RAND Corporation's definition of appropriateness as a measure of the balance between benefit and risk in a health care intervention [5]; and the application of Rothman's component cause model of causative factors in disease to genetic determinants of disease and the concept of penetrance [6].

1.4 This expanded framework, incorporating health quality measures, allows us to identify different components of clinical utility and to incorporate a consideration of ethical, legal, and social implications in the evaluation process.

2. Definition of a Genetic Test

2.1 We use as a starting point the definition of a genetic test and the distinction between an assay and a test as proposed by Zimmern and Kroese [2]. An assay may be defined as 'a method for determining the presence or quantity of a component' or 'a method to analyze or quantify a substance in a sample'. A genetic test is a laboratory assay that is used to identify:

1. a particular genotype or set of genotypes
2. for a particular disease
3. in a particular population
4. for a particular purpose

2.2 The reliability and accuracy of the assay (analytic validity) is primarily a matter of laboratory measurement. In the context of genetic testing it is a measure of the ability of the assay to measure the genotype of interest. The evaluation process requires precise definition of the genotype or set of genetic variants that comprise the assay. This consideration leads to a further distinction, between 'open ended' assays such as karyotype or mutation scanning across a gene, in which any abnormality is sought; and 'closed' assays, which specify in advance the spectrum of mutations or abnormalities the assay is designed to test. 'Open ended' assays may in turn be 'within a gene' or 'across the entire genome'.

2.3 The test relates the assay to a clinical situation. Context is all important. This requirement demands that the disorder or disease for which the assay is used is precisely defined; that the nature of the population on which the test will be applied is clearly set out, and most importantly that the purpose of the test is specified. The definition of the population is required because the positive predictive value of a test (PPV) of known sensitivity and specificity is related to the prevalence of the disorder within that population. In population screening tests, other epidemiological parameters such as the population attributable fraction may also be relevant.

2.4 We also point out that a disorder may be defined by genotype or by phenotype. If a genotypic definition is used, the assay is usually phenotypic in nature (for example, the sensitivity and specificity of Bethesda criteria can be evaluated against a genotypic definition of HNPCC as the finding of a pathogenic variant in *MLH1*, *MSH2* and *MSH6*); or, if genotypic in nature, must be based on a technology independent of the technology used to define the disorder (for example, a fluorescent *in situ* hybridization technique (FISH) might be used to test for a chromosomal disorder defined by conventional cytogenetic studies).

3. A Framework for the Evaluation of a Genetic Test

Introduction

3.1 The expanded framework for the evaluation of a genetic test builds on the ACCE framework and on our distinction between the assay and the test. The features of the framework are first, that it clarifies two separate components within the concept of clinical validity; second, that it identifies test purpose as a focus of test evaluation; and third, that it attempts to disentangle the various dimensions of clinical utility, using established measures of health care quality. These measures guide evaluation and provide a means for integrating ethical, legal, and social implications into the evaluation process. The framework leads to a three-step evaluation process:

1. evaluation of the assay
2. evaluation of clinical validity, including clinical test performance
3. evaluation of clinical utility, including test purpose and feasibility of test delivery

The Assay and the Test

3.2 The evaluation of the assay is primarily a matter of quality control and assessment. It is to be distinguished from general quality control and quality assurance measures for the laboratory in which the test is performed. It seeks to show that the test accurately and reliably measures the genotype of interest, that, for example, if it is intended to measure ten specified variants in the mythical gene CVA1 it does so with high sensitivity and specificity. It is synonymous with the measurement of analytical validity in the ACCE framework [1].

3.3 The evaluation of the test, by contrast, requires a definition of the disorder that the test seeks to diagnose or predict, a specification of the population to be tested and of the purpose of the test, followed then by an assessment of clinical validity and utility. Thus, in our example, we must ensure that proper scientific evidence links gene variants in CVA1 to increased risk of cerebrovascular accidents, that there has been formal evaluation of test performance in an appropriate clinical setting; and that the test performs adequately in accordance with the dimensions of clinical utility discussed below. The dimensions of test evaluation are set out in Table 1.

Test Purpose

3.4 The definition of a genetic test requires us to specify that testing is done for a particular purpose, for example, to identify a newborn who would benefit from early treatment of a genetic disorder, or to provide a family with information for reproductive decision-making. If the test is to achieve its purpose, it must be delivered in an appropriate way, in association with any relevant services or interventions.

Table 1: Dimensions of Test Evaluation

1. **Analytical validity of test**

2. **Clinical validity of test**

   (a) Evidence of gene-disease association

   (b) Measures of test performance

3. **Clinical utility of test**

   (a) Purpose of test

   (i) legitimacy
   (ii) efficacy
   (iii) effectiveness
   (iv) appropriateness

   (b) Feasibility of test delivery

   (i) acceptability
   (ii) efficiency
   (iii) optimality
   (iv) equity

3.5 The purposes for which genetic testing is done are listed here and discussed in detail in Annex 1. They are to:

1. Reduce morbidity or mortality

2. Provide information salient to the care of the patient or family members and/or

3. Assist the patient or family members with reproductive decision-making

Our discussion in Annex 1 addresses the reasons for specifying these purposes, as against the more familiar categories such as diagnosis, prediction or screening.

We suggest that the purpose of a newborn screening test, for example, is to reduce the morbidity or mortality of certain rare inborn errors of metabolism and that, therefore, its purpose will not be achieved unless children with positive test results are provided with appropriate interventions and follow-up care. These associated services will be of as much interest to patients and commissioners or funders of services as the technical aspects of the test itself. They include issues such as the need for counseling (either pre-test or post-test), the type of facility in which the test is carried out, interventions for individuals who test positive, patient satisfaction and many others. Established measures of health care quality (*legitimacy, efficacy, effectiveness, appropriateness*)

identify essential components in the evaluation of test purpose and will be discussed in Section 5.

Clinical Validity

3.6 Discussion of *clinical validity* frequently conflates two separate entities:

(1) the assessment of the link between genotype and disease (which in many cases is addressed through a systematic review of genetic association studies)

(2) the evaluation of test performance (by which we mean studies to determine characteristics such as sensitivity, specificity, PPV, NPV and likelihood ratios).

3.7 This confusion is particularly marked in regulatory circles where the terminology used is somewhat different and where the requirement that tests be 'clinically effective' may be interpreted in different ways. At one level, its meaning may be seen to require only prima facie scientific evidence of a link between the genetic variant (or variants) and the disorder; at another level, it could be interpreted to mean that evidence of a formal evaluation of test performance is needed.

3.8 The distinction allows these two separate aspects of the test to be discussed independently of each other. The establishment of gene-disease association is by and large an academic exercise undertaken by scientists and epidemiologists. Test performance is a specific activity requiring the establishment of a population with cases and controls and the application of both the test under evaluation and a reference (gold standard), in accordance with published standards for test evaluation. A serious policy issue (which will not be discussed further in this paper) is the lack of consensus about the responsibility, funding and mechanisms for generating these data.

3.9 We also return here to the distinction between 'closed' and 'open-ended' assays. An important element of the distinction is that it is only with *'closed' assays* that the standard 2 by 2 table can be used to assess the sensitivity and specificity of a genetic assay or test for detecting a particular health condition. With a closed assay, a direct and quantifiable comparison can be made between the outcome of the assay and the gold standard, allowing for accurate measurement of sensitivity, specificity, and positive and negative predictive value.

3.10 In contrast, accurate measures of sensitivity, specificity, and positive and negative predictive value cannot be calculated for open-ended assays. As a result, proxies must be used to estimate the clinical performance of an open-ended test. Microarray comparative genomic hybridization (CGH) offers an example. CGH represents a new technique for detecting sub-microscopic chromosomal abnormalities. It has potential utility in evaluating patients in whom a chromosomal abnormality is suspected (such as children with learning disabilities) when conventional chromosomal studies are normal. The test will identify a broad range of abnormalities, including some never detected before and some that are unlikely to be of clinical significance. In this setting, measures based on biological plausibility can be used to estimate the likelihood that a detected abnormality is clinically significant (eg, nature and location of the abnormality; whether parents with normal phenotypes also carry the genotype; whether the abnormality can be detected by alternative techniques such as FISH;

and whether similar chromosomal abnormalities have been described in normal populations) [8]. Using these parameters the test can be evaluated for its estimated diagnostic yield (proportion of those tested with a positive result) and false positive yield.

Clinical Utility

3.11 C*linical utility* is commonly understood to be a measure of the health care value provided by a test or technology. Thus, a genetic test is said to have clinical utility if it provides information that assists in the care of patients. Disagreements about the *clinical utility* of new genetic tests are common. Examples include debates about the appropriateness of screening for hemochromatosis [8.9], and the risks and benefits of expanded newborn screening panels [10,11]. Disagreements usually derive from different expectations about the purpose of genetic testing, and, for a given purpose, from differences of opinion about the evidence needed to establish that the test accomplishes its purpose.

3.12 We suggest that a more robust definition of *clinical utility* can be derived from:

(a) a consideration of the different purposes proposed for genetic testing

(b) the implications of the different ways in which a genotype contributes to the cause of disease and

(c) traditional measures of health care quality.

3.13 The manner in which a genotype contributes to disease is an important consideration when considering how tests are to be evaluated. We suggest that it is essential to distinguish tests for highly penetrant genotypes or inherited disorders from tests for complex disorders using genotypes with low penetrance that are in themselves not sufficient to give rise to the disease. When a genotype has incomplete penetrance, it is also important to distinguish between diseases that occur only when a genotype is present (such as acute intermittent porphyria) and diseases that may occur in the absence of a contributing genotype (such as venous thromboembolism). Genotypes that have high penetrance or are necessary causes of disease often provide salient health information to clinicians or patients and families even if interventions to improve outcome are lacking. However, low penetrance genotypes that increase risk for common diseases are unlikely to have utility unless they lead to interventions that reduce morbidity or mortality. This issue is discussed in further detail in Annex 2.

3.14 In addition to determining that a test and associated services can in fact achieve the intended purpose, policy makers must also consider whether delivery of these services is feasible within a given health care system. Thus, the dimensions of health quality may be subdivided into those that relate to the purpose of the test (legitimacy, efficacy, effectiveness, appropriateness), and those that relate to the feasibility of test delivery (acceptability, efficiency, optimality, equity).

3.15 The evaluation of feasibility includes two economic dimensions, optimality and efficiency. Economic evaluation may be applied to the assay, the test or the services

associated with test use. In our view it should be applied to the whole testing process, since delivery of the process represents the true cost of the 'test'. But even as between assay and test, different costs may be involved. At its simplest, a provision of the assay results alone cannot be equated with the provision and interpretation of a genetic test. Genomic technology is likely to reach a point where the generation of 'raw' results will be much less costly than the skilled manpower required to interpret them. Similarly, the cost of interventions for individuals who test positive may outweigh the cost of the test, particularly as we move towards genotyping as a guide to preventive strategies for common chronic disease.

3.16 Health care systems must also consider acceptability - the preferences of patients and their families regarding testing services - and equity. The equitable distribution of health care demands that tests which have been shown to be beneficial be available to all who have need of the service. Geographical differences in availability should be minimised. Equity also has an economic dimension. Equitable delivery demands that a health care system must consider the costs of ensuring equal services in all regions. The geographic distribution of relevant expertise and facilities may also present problems of efficiency.

3.17 The full and proper evaluation of a genetic test is necessarily a complex task and requires significant resources. It takes little reflection to see that time and resource constraints and logistical considerations dictate that it will not be possible for all tests (indeed most) to be subject to the full evaluative framework. Various levels of evaluation will have to be devised depending on the nature of the test, its purpose and the population in which it is to be carried out. Penetrance and the prevalence of the disorder or the genotype may also be significant factors in making this decision.

3.18 In addition, as we have noted, evaluation of clinical utility has two domains: evaluation of the test's potential to achieve its purpose, and evaluation of the feasibility of test delivery. Evaluation of these two domains will usually be sequential; and feasibility of test delivery may need to be assessed over time, after initial introduction of the test into clinical practice.

3.19 It is likely that most tests for rare disorders will require a modified and less stringent programme of evaluation than tests for common disorders or population screening. This is the case in part because penetrance is likely to be high and therefore association between a positive test and ultimate outcome is likely to be more predictable, and in part because the rarity of the test (and therefore the low expenditure for the test as a proportion of total expenditure for medical tests within the health care system) will mean that the cost of a full evaluation is unlikely to be justified. We suggest that development of criteria to determine those tests that should be subject to a full evaluation is a matter of priority.

4. Dimensions of Clinical Utility

Introduction

4.1 The framework that we suggest builds on the concept of the *quality* of a genetic test and associated health care services and sets out eight dimensions of quality by which a test should be evaluated. Our analysis derives from the work of Donabedian [3], who defines health care quality as the product of two factors; the science and technology of health care and the application of that science and technology in practice. He identifies seven dimensions by which quality can be evaluated. An eighth dimension is taken from work of the RAND corporation [5]. We review the application of each of these dimensions to genetic test evaluation. The first four relate to test purpose; the last four to the feasibility of test delivery.

Test Purpose

Legitimacy

4.2 *Legitimacy* refers to the conformity of a test to social preferences expressed in ethical principles, values, norms, mores, laws and regulations. In applying this quality measure to genetic testing, the different purposes of testing are relevant.

4.3 A test to reduce morbidity or mortality has a priori *legitimacy*, as long as it can be been shown to achieve its purpose and is introduced in conformity with governing laws and regulations. This is so because reduction of morbidity and mortality are the primary goals of health care. However, a test *proposed* for this purpose does not have legitimacy until there is evidence for its effectiveness. Legitimacy does not preclude a health care system or funder declining to offer or cover the test, if it is judged to be lacking other necessary qualities. In addition, the legitimacy of tests that inform treatment choices may be questioned in some cases. For example, variants of the hypothetical gene BALD1 might be good predictors of future baldness, and therefore might help to identify individuals who would benefit from topical minoxidil therapy to prevent hair loss. However, this genetic test might be judged an inappropriate use of health care resources because it addresses a normal physical state.

4.4 *Legitimacy* is also an important element in the evaluation of tests intended to provide health care information to the patient or family members. Most genetic tests currently used for this purpose provide information that is highly predictive and has obviously relevance to health care – for example, tests that identify a future risk of vision loss or diagnose a disabling condition such as Duchenne Muscular Dystrophy. The salience of the information to health care involves judgment, however, and disagreements may occur, particularly with tests that have limited predictive value. As an example, the value of *APOE* genotyping to predict risk of Alzheimer Disease, for which no specific treatment exists, has been debated.

4.5 *Legitimacy* is a key element in the evaluation of tests proposed for prenatal diagnosis or pre-implantation genetic diagnosis (PGD). As new opportunities for genetic diagnosis expand, policy makers will need to consider whether current standards

of practice adequately define the accepted limits for these services. Factors to be taken into account include the penetrance of genotypes; the severity of the associated conditions; and the degree to which thresholds for legitimate use should be different for PGD than for prenatal diagnosis.

4.6 The dimension of *legitimacy* is also reflected in regulatory procedures governing genetic tests. These include statutory regulation (oversight of laboratory quality, pre-market review of genetic tests, required post-market data collection); funding decisions; and professional guidelines for test use. Controversies over direct-to-consumer marketing or sale of genetic tests are an example of emerging questions about the legitimacy of genetic testing.

4.7 This quality dimension requires appropriate democratic procedures for determining legitimacy for different types of genetic tests. The procedures supporting determinations of legitimacy occur at several different points in policy development. The scope of statutory regulation, for example, is determined by legislative processes, ideally informed by testimony on technical issues and stakeholder concerns. Procedures are also needed to ensure appropriate societal input into health care funding decisions, practice guideline development, and other decision-making processes bearing on the use of genetic health care service.

Efficacy

4.8 *Efficacy* is the ability of the test (and any associated services) to bring about the intended purpose (often but not invariably improvements in health) when used under the most favorable circumstances. Sources of information about *efficacy* include research, experience and professional consensus [3]. The evidence required to establish the *efficacy* of a test will vary with test performance and purpose, and with the nature of the condition. In considering the available evidence, test evaluators need to define the outcomes of interest and the quality of evidence needed to evaluate them. The corollary is that it is not possible to make a statement about the *efficacy* (or effectiveness) of an intervention until and unless its purpose has been formally defined.

4.9 Some genetic tests allow a new clinical purpose to be addressed. For example, gene discovery may permit the development of a test to diagnose a rare untreatable single gene disease. The test provides a previously unavailable means for determining prognosis and offering prenatal diagnosis. In other instances the test provides a new way to achieve a purpose for which alternatives already exist. For example, *HFE* genotype testing provides an alternative to serum iron measures for the detection of people at risk for iron overload. When there is an alternative approach for achieving the test's purpose, evaluation of *efficacy* should include a comparison of the new genetic test with the existing alternative. At an even more basic level, evaluation of the efficacy of genetic tests may raise the question whether any form of DNA or molecular testing is required if clinical history or examination (the alternative) is able independently to provide an accurate diagnosis or assessment of risk.

4.10 Because the data available for genetic test evaluation is often very limited, two efforts are needed to improve the evaluation process:

(1) development of consensus on the type and quality of data required to justify test use, considering clinical test performance and the purpose(s) for which the test is proposed.

(2) development of strategies to generate needed data, such as funding allocations, identification of appropriate methodologies, and the potential creation of research consortia incorporating public-private partnerships as appropriate.

Effectiveness

4.11 *Effectiveness* is the degree to which attainable objectives (in most cases health benefits) are in fact attained under routine conditions. This quality measure emphasizes the point that a test benefit occurring in a research trial must ultimately be evaluated to determine whether the same benefit can also be delivered in the clinical setting. Key to this evaluation are factors that could result in differences between the research outcomes and the clinical outcomes – for example, lower laboratory quality; failure to refer appropriate patients; errors in test interpretation; failure to provide follow-up interventions; patient refusal of interventions. Each of these factors could identify a specific area of potential quality improvement. It is important in this assessment to consider logistical and other practical considerations. Even if a test has high efficacy, failure on the part of the delivery system to provide access or motivate the patient to seek services will render it relatively ineffective.

4.12 In the current environment, genetic tests often become clinically available after limited evaluation, usually with observational approaches subject to well established biases. Evidence for efficacy may therefore be weak. In this context, evaluation of *effectiveness* takes on a dual purpose: the delivery of the testing service is being evaluated, but in addition, the presumed efficacy is being evaluated by measurement of outcomes in practice.

4.13 A key element in the evaluation of *effectiveness* is a specification of the desired outcomes of testing, based on test purpose. As noted under efficacy, comparison with available alternatives is important. The outcomes of tests done to reduce morbidity and mortality may be easier to measure than outcomes of tests to provide salient information. Professional consensus may provide support for benefit with the latter group of tests, but evidence about the experience of individuals and families undergoing testing is also important. When a testing process is costly and the nature of the benefit subjective, formal studies (as opposed to expert and personal testimony) will be necessary to confirm the clinical utility of a test.

Appropriateness

4.14 The *appropriateness* of a health care service reflects the balance between the expected benefit and the expected negative consequences; the former should outweigh the latter by a "sufficiently wide margin" [5] that the service is worthwhile. Evaluation of *appropriateness* requires that both positive and negative outcomes of tests be defined and measured. Special attention to adverse consequences of testing may be needed when a test is proposed for population screening; when a test provides predictive information about a condition with serious social consequences; or when a test result leads to interventions that involve significant risk.

Summary

4.15 Taken together, the four dimensions of legitimacy, efficacy, effectiveness, and appropriateness provide a means to evaluate a test's purpose sufficiently to determine whether the test is appropriate for use in health care. This evaluation considers whether test use conforms to societal norms for health care, can achieve its intended purpose, and leads to sufficient net benefit to be adopted. While evaluation of these properties is essential for responsible decisions about test use, health care systems also have to consider a number of practical issues that bear on the feasibility of incorporating the test into the health care system. These issues are addressed by the remaining dimensions of health care quality, *acceptability, efficiency, optimality,* and *equity*.

Test Delivery

Acceptability

4.16 *Acceptability* refers to a test's conformity to the wishes, desires, and expectations of patients and their families. This quality measures points to the experience of patients and families as an important element in determining the services that should accompany a genetic test. The emphasis by advocates on the importance of appropriate and compassionate counseling [14] underscores the importance of this aspect of care.

4.17 Evaluation of *acceptability* requires inclusion of patients and families in the development and evaluation of genetic services. This participation may occur in several ways: (1) inclusion of representatives of the public and patient advocates in deliberations about tests and test services; (2) testimony from affected patients and families when guidelines for a specific test are developed or evaluated; (3) surveys or other studies to evaluate the outcomes of testing for patients and families and of the population tested.

4.18 As an example, an important element of the *acceptability* of prenatal diagnostic tests is the manner in which they are offered. Both non-directive counseling and adequate information have been identified as important components of the testing process. Disability advocates have noted that information provided to pregnant women about conditions like Down Syndrome is often limited or biased, potentially influencing decisions about prenatal diagnosis and pregnancy termination [12]. *Acceptability* of the testing service requires that couples have adequate, neutral sources of information, available in formats and venues that allow couples to choose what is best for them.

Economic Evaluation: Efficiency

4.19 *Efficiency* is the ability to lower the costs of care without diminishing benefits. This quality measure takes on particular importance when a service is beneficial but costly. As genomic technology evolves, and health care budgets are constrained, questions of *efficiency* will be increasingly important in the evaluation of genetic tests. A new test *assay* may be less costly, but does it provide the same positive and negative predictive value as the test currently in use? Or, the test may be more costly; does it provide sufficiently enhanced benefit to justify the cost? The delivery of genetic counseling services is also an area where questions of efficiency may arise, because the demand for counseling services may outstrip the availability

of trained personnel. In this context, one might ask whether better triage methods in the referring process, primary care education, substitution of other educational strategies (such as informational videos) for some of the counseling time, or related measures might result in more efficient use of counselor time without loss of benefit.

4.20 One way of achieving health care *efficiency* is to distribute care among different classes of patients so as to optimize outcomes ("distributional efficiency" [3]). Genetic testing represents a potential mechanism for achieving this efficiency, if it provides an effective means to utilize pharmaceutics, preventive or other services according to risk or predisposition. As with efficacy, consideration of alternatives is important: if there is a less expensive alternative, the test will have utility only if it provides added benefit. The marginal benefit of a newer technology over the marginal cost compared to the alternative is thus an important measure. From a population health perspective, the genetic test may have limited utility if the benefits of testing apply only to a small proportion of the population. For example, a cardiac prevention effort focused on individuals with familial hypercholesterolemia (FH) will provide less population benefit than a program focused on individuals with cholesterol levels in the top 5%, even though the individual benefit will be greater in the program focused on FH [13].

Economic Evaluation: Optimality

4.21 *Optimality* is the balancing of improvements in health against the costs of such improvements. This quality measure acknowledges the opportunity costs of medical innovation. It is particularly relevant for new genetic testing processes that involve expensive laboratory methods, counseling procedures, or interventions. Evaluation of optimality requires documentation of the costs of the testing process and any associated health care, as well as a quantifiable measure of benefit. A formal cost-effectiveness study may be appropriate for costly new testing services, or where the test provides limited benefits.

Equity

4.22 *Equity* refers to a test's conformity to the principle of just and fair distribution of health care and its benefits among members of the population. The question raised by this quality measure is whether equitable delivery the test is possible. If not (because of limited laboratory capacity or other resource or geographic constraints), it may be important to develop capacity before the test is introduced.

Summary

4.23 Taken together, these dimensions offer a means to assess whether the delivery of a test is feasible within a particular health care system. They incorporate both the manner in which testing services should be provided, and a practical consideration of the cost of the testing service and its priority relative to other health services.

Summary of Framework

4.24 Table 2 (next page) summarises the Framework that we propose.

Table 2: A Framework for Genetic Test Evaluation		
Domain	**Specific Element**	**Focus of evaluation**
Pre-evaluation definition	***Test Definition***	Precise definition of: Genetic variants to be assayed Disorder Population Purpose
Assay		
	Analytic validity	Sensitivity Specificity PPV, NPV
	Reliability and Reproducibility	Kappa
Clinical Validity		
	Gene-Disease Association	Primary research Systematic review Meta-analysis
	Clinical Test Performance	Sensitivity, Specificity PPV, NPV, LR+, LR-, ROC
Clinical Utility		
Test Purpose	***Legitimacy***	Conformity to the social preferences expressed in ethical principles, values, norms, mores, laws and regulations
	Efficacy	Potential of test and associated services to deliver health benefit
	Effectiveness	Actual delivery of health benefit in routine clinical setting
	Appropriateness	Expected health benefit exceeds expected negative cons quences by a sufficiently wide margin that the test is worth doing
Feasibility of Test Delivery	***Acceptability***	Conformity to the wishes, desires, and expectations of patients and their families
	Economic ***Efficiency*** ***Optimality***	 Ability to lower the costs of care without diminishing benefits Balancing improvements in health against costs of improvements
	Equity	Just and fair distribution of health care and its benefits among members of the population.

5. The Process of Evaluation

5.1 The evaluation of a test requires establishment of separate systems to:

(a) generate the data needed for the evaluation of test and associated services

(b) undertake the analysis and evaluation

(c) provide a policy response and clinical guidance

We take the view that the evidence for gene-disease associations will be found in original papers derived from epidemiological and clinical research and in systematic reviews and meta-analyses of the literature, and that their generation is primarily a matter for the scientific (including epidemiological) community.

5.2 We suggest that the largest gap at present is the lack of platforms and processes for generating the data needed to assess test performance in clinical settings. Systems and facilities for the systematic evaluation of genetic tests (and other complex molecular biomarkers) are poorly developed. Studies of the diagnostic accuracy of tests (done primarily in the context of biochemical, radiographic or histological tests) show that test evaluation is frequently subject to a variety of design related biases and flaws [15,16] Methodological standards exist for both diagnostic test evaluation and for their systematic review [17], but the evidence suggests that in most instances these have not been followed [16]. From a logistical standpoint, it will not be possible to evaluate formally every new test or biomarker. For this reason, criteria will be needed to ensure those tests with the greatest clinical or economic implications are properly assessed. We believe that these criteria should be developed as an urgent priority.

5.3 There has been also much discussion as to whether the responsibility for the generation of such data should be placed entirely on the commercial sector or whether government and other public agencies should play a role. Whatever conclusion is reached in relation to the generation of data concerning test performance, there appears to be a developing consensus that the analysis and evaluation of those data should be carried out independently, and without the direct involvement of those responsible for the development, provision or marketing of such tests.

5.4 We believe that such evaluation should take into account not just the technical characteristics of the test but also the entire spectrum of quality standards that we discuss in this paper, some of which relate to the entire testing service or test process. We also suggest that the process of analysis and evaluation should be separated from those institutions and organisations responsible for making policy and establishing guidelines for how tests should be used in clinical practice. Our framework generates a set of questions for evaluators, related to each component of test evaluation. These are summarized in Table 3 (following page).

6. Conclusion

6.1 This paper makes the case for an enhanced schema for the evaluation of genetic tests. Our proposals build on the ACCE framework. They include:

(a) the requirement to distinguish the *assay* from the *test*; and tests for highly penetrant inherited disorders from tests for complex disorders involving low penetrance gene variants

(b) the need to appreciate that two separate and distinct concepts are embedded within the dimension of *clinical validity*: the assessment of the link between genotype and disease, and the assessment of clinical test performance

(c) the importance of defining the *purpose* of the test and of relating concepts of *efficacy* and *effectiveness* to the stated purpose

(d) much greater attention to the assessment of clinical *utility,* using concepts developed by Donabedian and the Rand corporation for quality assurance in health care

6.2 We suggest that this enhanced framework will provide a more robust and rational approach to genetic test evaluation, not least because it gives much greater emphasis to factors that move beyond the epidemiological and technical to embrace those important to the patients and their families and to the entire process of test delivery, including associated health care services.

7. References

1. Haddow J, Palomaki G. ACCE: a model process for evaluating data on emerging genetic tests, in Khoury M, Little J, Burke W, eds. Human Genome Epidemiology, Oxford University Press, Oxford 2004, pp. 217-233.

2. Zimmern R, Kroese M, The evaluation of genetic tests. Public Health (oxf.) 2007 Sep;29(3):246-50.

3. Donabedian A. An Introduction to Quality Assurance in Health Care. Oxford University Press, Oxford, 2003.

4. Hopkins A. Measuring the Quality of Medical Care. Royal College of Physicians, London, 1990.

5. Kahn KL, Kosecoff J, Chassin MR et al. Measuring the clinical appropriateness of the use of a procedure; can we do it? Med Care 1988; 26:415-22.

6. Rothman K. Modern Epidemiology. Little, Brown. Boston. 1986, pp.7-23.

7. Subramonia-Tyer S, Sanderson S, Sagoo G, Higgins J, Burton H, Zimmern R, Kroese M, Brice P, Shaw-Smith C. Array based comparative genomic hybridisation for investigating chromosomal abnormalities in patients with learning disability: systematic review meta-analysis of diagnostic and false positive yields. Genetics in Medicine 2007; 9:74-79.

8. McDonnell SM Parrish RG. Hereditary hemochromatosis and its elusive natural history. Arch Intern Med. 2003 Nov 10;163:2421-3.

9. Dubois, Kowdley KP. The importance of screening for hemochromatosis. Arch Intern Med. 2003 Nov 10;163:2424-6.

10. Botkin J, Clayton EW, Fost NC *et al*. Newborn screening technology: proceed with caution. Pediatrics. 2006 May;117:1793-9.

11. Howell RR. We need expanded newborn screening. 2006 May;117:1800-5.

12. Parens E, Asch A. Disability rights critique of prenatal genetic testing: reflections and recommendations. Ment Retard Dev Disabil Res Rev. 2003;9:40-7.

13. Grosse SD, Khoury MJ. What is the clinical utility of genetic testing? Genet Med 2006;8:448-450.

14. For example, comments by Alastair Kent at an OECD meeting, Manchester, UK, June 27, 2006.

15. Lijmer JG, Mol BW, Heisterkamp S *et al*. Emprirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999; 282:1061-1066.

16. Tatsioni A, Zarin DA, Aronson N *et al*. Challenges in systematic reviews of diagnostic technologies. Ann Intern Med 2005; 142:1048-1055.

17. Bossuyt PM, Reitsma JB, Bruns DE *et al*. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. BMJ 2003; 326: 41-44.

Annex 1

Purpose of genetic tests

1. Genetic tests have a range of different purposes in health care. In addition, a particular test *assay* may be used for multiple purposes. Test purpose must be defined before test evaluation can occur, because the utility of a test is defined in terms of its purpose or objective. Based on current practice, we propose that all genetic tests have one or more of the following purposes:

    1. Reduction in morbidity or mortality

    2. Provision of information salient to the health care of the patient or family members

    3. Assistance with reproductive decision-making for patient or family members

2. We advocate that this classification provides a better basis for test evaluation than the more usual designation of test purpose in general terms such as diagnosis, predictive testing, susceptibility testing or screening. We take the view that these terms serve only as *intermediate purposes* since they do not get to the heart of why we undertake a particular test. A test result, whether for diagnosis or prediction or as a result of a screening programme, is not an end in itself. We seek, for example, to make a diagnosis because it allows us to undertake appropriate therapeutic or preventive interventions that might lead to reductions in morbidity or mortality, to inform the patient (or family) about a health risk that will influence the care of the patient, or to provide prenatal diagnosis. We believe that an adequate evaluation of purpose requires the more precise and more outcome-orientated taxonomy described here.

3. We also point out at this stage that the legitimacy and appropriateness of some genetic test purposes – in particular the use of genetic testing to provide salient health information or for reproductive decision-making - is influenced by penetrance and other considerations of genetic causality. The use of an innovative genetic test in clinical practice also raises another issue of legitimacy, the distinction between clinical care and research. These important aspects of genetic testing are further discussed in Annex 2 and Annex 3.

Testing to reduce morbidity or mortality

4. The most straightforward claim for clinical utility arises when a test is proposed as a means to reduce morbidity or mortality. Cogent examples include the identification of infants with phenylketonuria (PKU) by newborn screening, so that the appropriate diet can be used to prevent mental retardation; *RET* mutation testing to identify persons at risk for medullary thyroid cancer who will benefit from prophylactic thyroidectomy and monitoring for pheochromocytoma and hyperparathyroidism; and testing for HNPCC-associated mutations to identify patients who will benefit from early and aggressive

colon cancer screening. As these examples demonstrate, reducing morbidity or mortality requires both an appropriate testing process and an effective intervention for those found to have the risk-associated genotype.

5. Diagnostic, predictive, and screening tests can be used for this purpose, as

Annex 1 Table: Genetic tests used to prevent morbidity or mortality

Test	Person tested	Type of test
Urinary prophobilinogen level; sequencing of HMBS gene [1]	30 year old woman with recurrent abdominal pain, nausea and vomiting and signs of peripheral neuropathy	Diagnostic: if positive, acute intermittent porphyria (AIP) is diagnosed, leading to specific management of diet and medications to prevent or minimize future attacks.
Test for HMBS mutation [1]	First degree relative of patient with AIP and known HMBS mutation	Predictive: if positive, patient is at risk to have symptoms of AIP, and will benefit from avoiding known precipitants of AIP symptoms
Test for mutation in the MYOC [2]	21 year old man with family history of early onset glaucoma associated with MYOC mutation	Predictive: if positive, increased risk of glaucoma is identified; early screening is appropriate
Test for phenylalanine level [3]	Newborn	Population screening: If positive, leads to confirmatory testing to identify PKU, for purpose of offering diet to prevent development of mental retardation
Measurement of iron in liver biopsy specimen [4]	40 year old man with hepatomegaly, elevated liver function tests and serum ferritin >750	Diagnostic: If positive, confirms hemochromatosis, leading to phlebotomy treatment

6. The key questions for this purpose are whether an effective intervention is available and whether the use of the test to identify candidates for the intervention results in an acceptable ratio of benefit to harm (the concept of *appropriateness)*. Penetrance is not itself a crucial factor in determining whether a test can fulfill the purpose of reducing morbidity or mortality. Highly penetrant genotypes have a higher predictive value, which may increase the benefit of an intervention that reduces morbidity or mortality. However, a genotype with low penetrance may provide an important health benefit if it leads to the use of a safe and effective therapy; for example, it is likely that only 20% to 25% of infants who screen positive for medium chain acyl-CoA dehydrogenase

deficiency (MCADD) would suffer serious disability or premature death if untreated [5], but simple dietary measures that are safe for all individuals with MCADD can prevent these adverse outcomes.

7. When a genotype is a necessary cause of disease, a positive test result may provide essential information about appropriate interventions. For example, people with acute intermittent porphyria (AIP) are recommended to avoid certain exposures that are safe for others; and a person with hemochromatosis requires a specific treatment, periodic phlebotomy, that could be harmful in people without the disorder.

8. When the intervention is not genotype-specific, the value of testing requires careful consideration. For example, some gene variants identify people with moderately increased risk for coronary heart disease. The appropriate management of such individuals includes the same coronary prevention measures (diet, avoidance of smoking, exercise) suggested for the general population. In contrast to AIP and hemochromatosis, identification of genotype is not necessary to select the interventions that will reduce morbidity or mortality; however, the genotype may have clinical utility if it can be shown to motivate a higher level of compliance with the recommended lifestyle changes. Whether such knowledge does lead to greater compliance is as yet an unresolved question and the focus of much research activity.

Testing to provide salient information for the health care of the patient or family members

9. A test may provide medical value even if a reduction in morbidity or mortality is not possible. For example, a diagnosis of Duchenne Muscular Dystrophy provides prognostic information, and allows for planning of supportive and palliative care based on knowledge of the natural history of the disorder. Similarly, vision loss in the retinal dystrophies cannot be prevented or delayed, but a diagnosis provides prognostic information and guidance regarding the appropriate frequency of ophthalmic examinations. In some cases, the benefits of such care may be sufficient to qualify as reductions in morbidity. Even when this claim cannot be made, if the test provides a diagnosis or prognostic information, care is likely to be more efficient, and unnecessary work-up is avoided. Once a diagnosis is made within a family, testing may provide similar benefits to other family members. A negative genetic test result can also be helpful if it rules out a genetic condition, thus removing the risk and the need for any associated follow-up.

10. When a genotype has incomplete penetrance, prognostic information is less certain. However, a genetic test result may still have the potential to improve the process of care. A patient who is known to have a genotype associated with a variable degree of vision loss will benefit from careful monitoring of vision, even if the natural history of the disease is uncertain and cannot be changed. Identification of a genetic diagnosis may also simplify care by ruling out other considerations. For example, a diagnosis of Gilbert disease explains episodic elevations of bilirubin and allows the patient to avoid further work-up.

11. Knowledge of a specific genetic diagnosis may also contribute to patient or family well-being. In an interview study [5], parents of disabled children diagnosed with specific genetic conditions identified several benefits of the diagnosis, including relief of parental guilt; resolution of uncertainty; a clearer understanding of the child's future

needs; the potential to identify others in the same situation, for mutual support; and in some cases help justifying services. The knowledge itself may be viewed as a benefit. A genetic diagnosis may explain a rare, unexpected and painful event – such as a child who is blind or deaf, or a fatal cancer in a young adult - and clinicians note that families often want an explanation for why such events happened.

12. The value of the information varies with different conditions and tests. The diagnosis of an ophthalmologic disorder that will lead to severe visual impairment may help to plan appropriately for a child's education and career. Similarly, knowledge of future risk for disabling neurological disease could allow a person to do appropriate financial planning and help family members to prepare for the impact of the disorder. Psychological benefits have been postulated - for example, that the identification of a genetic cause for obesity or mental illness might relieve guilt or reduce stigma.

13. Although information can be viewed as a benefit if it explains an event or allows planning and support, it may also be a source of grief and pain. Knowing that a child will inevitably lose vision, or worse, experience progressive neurological degeneration and early death (as is the case when juvenile Batten Disease is diagnosed) may be extraordinarily burdensome to parents. The information may be essential for understanding a medical problem, and may assist individuals or families to adjust to a tragic situation, but may not be perceived as a contributor to their well-being. Testing done in this context needs to take into account factors such as timing of testing, and the pre- and post-test counseling, that may increase the effectiveness and acceptability of the testing process.

Testing to assist reproductive decision-making

14. Genetic testing can inform individuals or couples about their risk to have a child with a genetic disease. This use of genetic testing has unique implications with respect to legitimacy and acceptability.

15. In many developed countries, certain carrier and prenatal tests are offered routinely to pregnant women. These include tests to identify an increased risk of Down syndrome and tests to identify carriers for autosomal recessive diseases among populations with an increased prevalence, such as Tay-Sachs disease in Ashkenazi Jewish populations, and beta-thalassemia in Mediterranean communities. If testing leads to the prenatal diagnosis of an affected child, pregnancy termination can be offered. However, this option is not universally available; for example, pregnancy termination is not an option in Chile or many mid-Eastern countries.

16. Carrier and prenatal testing are also offered to families after the diagnosis of a child with a severe single gene disorder, such as Duchenne Muscular Dystrophy, cystic fibrosis, or X-linked retinitis pigmentosa. Usually, testing is offered in order to provide the choice of pregnancy termination. In some instances, families wish to pursue prenatal diagnosis in order to prepare for the birth of a child with a genetic disease.

17. The acceptability of prenatal diagnosis and selective pregnancy termination vary widely among individuals in countries where these services are available. For this reason, medical genetics practice calls for non-directive counseling when carrier and prenatal diagnostic tests are offered. Health care systems must also consider what tests can legitimately be offered for this purpose.

18. In the past two decades, an additional reproductive option has become available for couples at risk to have a child with a genetic disease: pre-implantation genetic diagnosis (PGD). For some couples, selection of an embryo for its genetic characteristics is morally preferable to selective pregnancy termination. From a health care systems perspective, PGD poses two challenges. The first is the cost and associated risk of this procedure, which require that it be offered selectively. The second is the need to establish an appropriate decision-making process to determine which conditions merit this prevention strategy.

Family-based testing

19. For many of the examples cited, family-based detection of genetic risk is a method to fully achieve the goals of testing process. For example, after the diagnosis of acute intermittent porphyria, first degree relatives are offered testing, in order to identify affected relatives who would benefit from preventive measures (Annex Table 1). Reproductive decision making may also be a reason for family-based testing: cystic fibrosis (CF) carrier testing is offered to first degree relatives of parents of a child with CF, to determine whether they also are at risk to have a child with CF.

20. As these examples illustrate, family-based detection is not an independent purpose of genetic testing, but rather is linked to achievement of one of the other purposes of testing. In highly penetrant single gene conditions, family-based testing is an efficient means to identify persons at risk, because the prior probability of a positive test result is high.

Intermediate Purposes

21. Notwithstanding our preference for using ultimate purpose as the primary designation of test evaluation the distinctions between diagnosis, predictive testing and screening remain important. These distinctions determine how services are offered, and have implications for potential benefits and harms of testing. A test is *diagnostic* when it seeks to explain symptoms for which a patient has sought help. Tests are chosen based on clinical presentation, with the goal of defining the nature of the problem and when possible to guide treatment. The presence of specific symptoms influences the prior probability of a positive test result.

22. A test is *predictive* when it seeks to identify a health risk in an asymptomatic patient. Predictive tests may in turn be subdivided into (a) pre-symptomatic or (b) susceptibility tests. If the penetrance of the genotype identified by the test is high, the test is considered *pre-symptomatic* – that is, there is a high likelihood, in some cases a certainty, that the patient will ultimately develop symptoms of the disease.

23. *Pre-symptomatic testing* is typically done within a family context. Sometimes, pre-symptomatic testing is offered on the basis of family history reported by the patient: testing for Huntington Disease or MEN 2 may be offered on this basis. Usually, however, testing is initiated in an affected family member. This approach takes into account the cost and incomplete sensitivity of most genetic tests: if a specific causative genotype is identified in the affected family member, a less expensive test with higher positive

and negative predictive value (testing for the specific causative genotype found in the affected family member) can be offered to relatives. If no genotype is identified in the affected relative, testing is not informative in unaffected family members.

24. If the penetrance of the genotype is low, the test identifies a genetic susceptibility; in other words it offers an estimation of the risk or susceptibility of developing a particular disease. *Genetic susceptibility tests* might be offered on the basis of a patient's concern, or family history, or other clinical history suggesting risk. For example, some expert groups have recommended thrombophilia testing for women who have experienced recurrent pregnancy loss, because thrombophilia is a known risk factor for this obstetric problem.

25. *Screening* refers to testing done on the entire population, or in some subset of the population where there is evidence to suggest that risk might be increased. These populations are usually defined demographically. The other defining feature is that in *screening* (as distinct from *testing*) the test is not initiated by a patient who seeks advice to alleviate anxiety or to solve a problem, but by the health care system or the state, which has identified the person as someone who might benefit from the information provided by the screening test. Sometimes, the group offered screening is a sub-population with an *a priori* higher risk (as when women over age 50 are offered mammography screening); at other times, screening is offered to the population at large. Newborn screening represents an example of a genetic screening program. Another example is the use of family history as a screening tool in primary care practice to identify people at increased risk for cancer, diabetes or cardiovascular disease.

26. *Diagnostic*, *predictive*, and *screening* tests can all serve the different purposes of genetic tests. The intermediate designation of a test as diagnostic, predictive, or used in screening does not therefore fully clarify its purpose. However, these distinctions help to define the population in which testing will be done, and have implications for determining benefits and risks. Concerns about adverse labelling as a result of a positive test result are generally lower for diagnostic than for predictive tests. Concerns about false positive results are particularly high for screening tests, because screening tests are done in asymptomatic persons without health concerns and because prevalence of the disease for which screening is done is generally low.

Annex 1 References

1. Sassa S. Acute intermittent porphyria (GeneReview) 2005, www.genetests.org, accessed July 28, 2006.

2. Libby RT, Goould DB, Anderson MG, John SWM. Complex genetics of glaucoma susceptibility. Annu Rev Genomics Hum Genet 2005; 6:15-44.

3. National Institutes of Health Consensus Development Conference Statement: phenylketonuria: screening and management, October 16-18, 2000. Pediatrics 2001 Oct; 108:972-82.

4. Powell LW, George DK, McDonnell SM, Kowdley KV. Diagnosis of hemochromatosis. Ann Intern Med 1998; 129:925-931.

5. Grosse S, Khoury MJ, Greene CL, Crider KS, Pollitt RJ. The epidemiology of medium chain acyl-CoA dehydrogenase deficiency: an update. Genet Med 2006; 8:205-212.

6. Gogarty B. Parents as partners. Cambridge Genetics Knowledge Park 2005. Report can be accessed at www.cgkp.org.uk.

Annex 2

Penetrance and Genetic Causality

1. Some genetic disorders demonstrate complete penetrance – that is, all people with the genotype have the disorder. An example is trisomy 18: all people with this disorder have developmental delay and dysmorphisms [1]. In epidemiological terms, the genotype represents the minimally sufficient condition to cause the disease [2]. Other conditions have complete or near-complete penetrance under conditions that are universal or very common; examples include phenylketonuria (which requires both genotype and a normal diet) [3] and Huntington Disease (which requires both genotype and sufficient longevity to manifest the disease) [4]. In addition, some genetic disorders have a very high, if not complete, penetrance; for example, familial adenomatous polyposis is estimated to confer a 93% risk of colorectal cancer by age 50 [5] and lifetime risk is assumed to approach 100%.

2. Other genotypes have incomplete penetrance. The *HFE* genotype C282Y/C282Y results in an increased risk of iron overload and associated complications, such as cirrhosis and diabetes. The penetrance of the genotype is still not precisely defined, but current estimates suggest that only 1% to 9% of individuals with this genotype will experience clinical symptoms [6,7]. Other gene variants confer small or moderately increased risks for cancer, cardiovascular disease, diabetes and other common complex diseases.

3. Another important consideration is whether the genotype is necessary for disease to occur. Acute intermittent porphyria (AIP) is an autosomal dominant disorder that results in acute episodes of abdominal pain and other gastrointestinal symptoms, and may also include peripheral neuropathy and psychiatric symptoms [8]. AIP represents a unique clinical phenotype, and, so far as is known, occurs only when a person has a mutation in the erythrocyte hydroxymethylbilane synthase (*HMBS*) gene. However, not all people who have a disease-causing *HMBS* gene mutation manifest symptoms. Thus, this genotype is *necessary but not sufficient* to cause disease. The reason for incomplete penetrance is presumed to be the variable exposure of at-risk persons to precipitating agents, such as alcohol, steroids, barbiturates, sulfa-containing drugs and other chemicals, and smoking, or precipitating stresses such as reduced caloric intake, infections, surgery, and long air travel. Overall, penetrance is estimated to range from 10% to 52% [8]. Despite this uncertainty, the test can provide useful diagnostic and predictive information, because the AIP phenotype occurs only in the presence of a specific genotype. As a result, a positive genetic test result can diagnose persons presenting with symptoms compatible with AIP, and can provide useful information to relatives about the advisability of avoiding certain exposures.

4. By contrast, the Factor V Leiden (FVL) genotype increases risk for venous thrombosis (VTE), with an estimated lifetime risk of 10% to 20% [9], but is not necessary for VTE to occur. In epidemiological terms, this genotype represents a causal component that is neither necessary nor sufficient to cause the disease in question. Genetic risk factors of this kind provide a means for identifying a group at increased risk. As with other risk factors, clinical utility will depend on whether identification of the risk group leads to opportunities to improve heath care.

5. We suggest that genotype *penetrance* and genotype as a *necessary or non-necessary cause* of disease are important considerations in assessing the legitimacy and appropriateness of a genetic test. The high predictive value that results from complete or near-complete penetrance increases the likelihood that a test result will improve the process of health care or provide useful information to families. When a genotype has incomplete penetrance, its clinical utility is likely to be greater if the genotype is a necessary cause of disease, as compared to genotypes that merely increase risk but are not invariably present in people with the disease.

6. In addition, complete or near-complete penetrance is an important factor in considering the legitimacy of selective pregnancy termination. While genotypes with significantly reduced penetrance may provide clinically useful information, they are not likely to be viewed as legitimate tests for prenatal diagnosis.

Annex 2 References

1. Hunter AGW. Medical genetics 2: The diagnostic approach to the child with dysmorphic signs. CMAJ 2002; 167: 367-72.

2. Rothman K. Modern Epidemiology. Little, Brown. Boston. 1986, pp.7-23.

3. National Institutes of Health Consensus Development Conference Statement: phenylketonuria: screening and management, October 16-18, 2000. Pediatrics 2001 Oct; 108:972-82.

4. Haigh B, Huq M, Hayden MR. Huntington disease (GeneReview) 2005. www.genetests.org, accessed July 28, 2006.

5. Solomon C, Burt RW. APC-associated polyposis conditions. (GeneReview) 2005. www.genetests.org, accessed July 28, 2006.

6. O'Neil J, Powell L. Clinical aspects of hemochromatosis. Semin Liver Dis 2005 Nov; 25(4):381-91

7. Beutler E, Felitti VJ, Koziol JA *et al*. Penetrance of 845G--› A (C282Y) HFE hereditary haemochromatosis mutation in the USA. Lancet 2002; 359:211-218.

8. Sassa S. Acute intermittent porphyria (GeneReview) 2005, www.genetests.org, accessed July 28, 2006.

9. Juul K, Tybjaerg-Hansen A, Schnohr P, Nordestgaard BG. Factor V Leiden and the risk for venous thromboembolism in the adult Danish population. Ann Intern Med 2004; 140:330-7.

Annex 3

Distinguishing between research and innovative clinical care

1. Clinicians sometimes use genetic testing to evaluate puzzling clinical findings which could represent either an atypical presentation of a known genetic condition or a previously undescribed genetic condition. As open-ended genetic tests increase in resolution, they may be applied for this purpose to patients with clinical findings suggesting a genetic etiology. Increasingly precise measures of chromosomal abnormalities are a case in point. Among children with learning disabilities, microarray CGH is estimated to identify an abnormality in 13% of children in whom conventional cytogenetic analysis was normal [1]. In addition to providing families with an explanation for the child's problem, these test results may help clinicians to further classify subsets of patients with learning disabilities. Ultimately, prognostic information and guidance for clinical management may emerge, particularly if information about clinical findings in patients with different abnormalities is captured in databanks such as DECIPHER [2].

2. This use of testing raises the question of the dividing line between clinical care and research. The issue is not specific to genetics but may arise more frequently in genetics because of the current rapid development of genomic technology. In this setting, clinicians must distinguish between the use of an innovative test to benefit the patient and use of the same test to produce generalisable knowledge about patient care. The former is a legitimate part of clinical practice; the latter constitutes research and should include appropriate informed consent and study design. When tests are used as part of patient care, they may still also contribute to growing knowledge about genotype-phenotype correlations or outcomes of therapy, if clinicians are able to contribute to data gathering efforts such as DECIPHER that permit the collation and analysis of information gathered during clinical care.

3. In genetic test evaluation, this issue is relevant when a new genetic test is proposed for the purpose of improving the process of health care or providing useful health information to families. Legitimate use of a test in clinical care for these purposes requires that it meets the same performance and quality standards required of tests already in use for these purposes.

Annex 3 References

1. Sanderson S, Zimmern R, Higgins J *et al*. Challenges for evaluating genetic tests: lessons learned from an evaluation of array-based comparative genomic hybridization for investigating chromosomal rearrangements in patients with learning disability, in preparation.

2. DatabasE of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources (DECIPER). www.sanger.ac.uk/PostGenomics/decipher/, accessed July 28, 2006.