



The evaluation of diagnostic laboratory tests and complex biomarkers

**Summary of a Diagnostic Summit
14 - 15 January 2008**



**Professor Peter Furness
Dr Ron Zimmern
Dr Caroline Wright
Dr Maria Adams**

For copies of this report please contact:

PHG Foundation
Strangeways Laboratory
Worts Causeway
CAMBRIDGE
CB1 8RN

www.phgfoundation.org

Royal College of Pathologists
2 Carlton House Terrace
LONDON
SW1Y 5AF

www.rcpath.org

Executive Summary	4
1 Introduction	5
2 Summit Presentations	7
2.1 Introduction: a pathologist’s perspective	7
2.2 Introduction: a public health perspective	8
2.3 Evaluating new laboratory investigations in the context of NHS commissioning	9
2.4 Developing Labtests Online	10
2.5 Evaluation of genetic tests: the experience of the UKGTN	11
2.6 What information does the health service need on diagnostics, and how should it be delivered?	12
2.7 Additional short presentations	12
3 Discussion	14
3.1 Database of tests	14
3.2 Generation of data for test evaluation	15
3.3 Test evaluation	16
3.4 Regulatory framework	17
4 Recommendations	19
5 Appendices	20
Appendix 1 Glossary of terms	20
Appendix 2 Summit programme	22
Appendix 3 Summit participants	23
Appendix 4 Speaker biographies	25
Appendix 5 List of papers tabled	26
6 References	27

Executive Summary

A Diagnostics Summit was held at the Genome Campus in Hinxton, Cambridge on the 14-15th January 2008, organised by the Royal College of Pathologists (RCPATH) and the Foundation for Genomics and Population Health (PHG Foundation). The meeting consisted of invited participants, primarily from the UK, with expertise in diagnostic test evaluation, health technology assessment, healthcare policy and research.

The objective of the meeting was to agree a set of recommendations for the evaluation and regulation of clinical laboratory tests and complex biomarkers. The delegates to the summit were of the view that it was now necessary for government and the Department of Health to give greater priority to these matters. They unanimously agreed the 10 recommendations detailed in this report.

The delegates suggested that a new body should be established to ensure the evaluation of laboratory diagnostic tests. They also called for the creation of a publicly available database of new and existing laboratory tests which should contain evidence of clinical performance, as far as that evidence is available. Where evidence is missing, particularly evidence of clinical validity and clinical utility, this should be explicitly stated. Policy makers and all stakeholders should be encouraged to address issues around gathering the necessary evidence for the clinical evaluation of new and complex biomarkers, and the involvement of industry. An independent expert body should be responsible for evaluating the evidence for test performance and for making recommendations about appropriate clinical use. Such an evaluation could be used to enable tests to be placed on a 'diagnostics formulary' which identifies those tests considered to have clinical utility, as well as describing the ways in which the tests could be used. Direct-to-consumer testing was not discussed in detail, but there was agreement that this was an issue that would require further consideration.

Delegates further recommended that commissioners and health care professionals should be encouraged to use only those tests where appropriate evidence of clinical performance exists. Statutory regulators should be empowered to require that evidence (or lack of) relating to test performance be placed in the public domain and should call for a more responsive and proportionate risk assessment to ensure patient safety.

Such an evidence base would not only assist health service professionals, providers and patients, but would also provide much-needed clarity for commercial organisations and academic researchers who wish to bring their innovations into NHS use.

1 Introduction

The rapid and accurate diagnosis of a patient's condition is an essential part of clinical management, and laboratory tests are a vital part of this process. Testing is also used to tailor individual treatment plans according to need, to monitor disease progression, to stratify risk, to inform prognosis and for population screening programs. Together with other diagnostic technologies laboratory tests provide an essential component of the patient's journey within the health care system.

In the UK, around 1 billion laboratory tests are performed each year. NHS laboratories have sophisticated systems to ensure the analytical accuracy of the tests, yet no system is in place to ensure the clinical effectiveness and utility of individual tests. This present situation is analogous to having a pharmaceutical industry with tight control of the chemical purity of drugs, but with no formal requirement for evidence that a drug benefits patients.

A proper system for the evaluation of clinical laboratory tests would be of enormous benefit to a number of key stakeholders, including clinicians who must decide how to diagnose and treat a patient, pathologists who implement tests and interpret results, commissioners of health services who have to decide which tests should be available on the NHS, researchers who seek funding to identify biomarker-disease associations, anyone seeking to audit the quality of patient investigation, and companies whose role is to develop and bring new tests to market.

A 2006 report from the Royal College of Pathologists (RCPATH)¹ noted that, in light of the increasing complexity and rate at which new diagnostic laboratory tests are becoming available, historical approaches to evaluation of the clinical relevance and utility of such tests are becoming increasingly inadequate. The report recommended that a working group be established to discuss how best to develop an authoritative mechanism for the evaluation of new laboratory investigations.

In the same year, the Cooksey Report² highlighted the gap between “translating ideas from basic and clinical research into the development of new products and approaches to treatment of disease and illness, as well as implementing those new products and approaches into clinical practice”. A recent paper by Khoury et al.³ identified four stages in the translation continuum:

- T₁ – Discovery of biomarker/disease association to candidate health application
- T₂ – Health application to evidence-based practice guidelines
- T₃ – Practice guidelines to health practice
- T₄ – Practice to population health impact

The association of a disease with a biomarker – a characteristic that can be objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention – is currently the focus of the majority of research funding (T₁). However, the second stage of translation (T₂) – specifically, the need for evaluation of the clinical validity and usefulness of diagnostic tests – is often ignored, resulting in a limited evidence base for investing in and using a particular test⁴. Despite the critical importance of evaluation to the provision of good healthcare, the field of diagnostics suffers from a lack of coherence, interest and investment in this area.

The PHG Foundation has taken specific interest in the policy issues concerned with the evaluation and regulation of genetic tests⁵, and has been closely involved in developing processes for so doing within the United Kingdom Genetic Testing Network (UKGTN). The parameters required for proper evaluation of diagnostic tests are known⁶ and a full framework for evaluating genetic tests has been developed^{7:8}, refined⁹ and implemented¹⁰. One important insight from this work has been that, before a test can be properly evaluated, its purpose must be clearly defined¹¹ in addition to the particular disease and the exact population group for which the test is intended¹². Where a laboratory assay is used in a variety of clinical settings, taking account of these variables adds to the complexity of test evaluation, but cannot safely be ignored.

The ACCE model provides a theoretical framework for the evaluation of a test. The key features of this model are:

- Measurement of the **analytical validity** of the assay, i.e. the accuracy and precision with which a particular biomarker is identified by the test
- Evaluation of the **clinical validity** of the test, i.e. the accuracy with which a test identifies or predicts a patient's clinical status
- Assessment of the **clinical utility** of the test, i.e. assessment of the risks and benefits, such as cost or patient outcome, resulting from using the test
- Consideration of any **ethical, legal or social implications** (ELSI) resulting from using the test

The first of these is usually well covered by laboratory quality assurance procedures. The second is sometimes assessed in research publications that report sensitivity, specificity, predictive values and related parameters. But the third and fourth items are often not formally evaluated at all, despite being key to determining whether or not the test actually produces a benefit.

Recognising that this situation is problematic, the RCPATH and the PHG Foundation co-hosted a Diagnostics Summit, in which key stakeholders and experts were invited to discuss a mechanism by which the systematic evaluation of new and existing clinical laboratory tests could be implemented.

2 Summit Presentations

2.1 Introduction: a pathologist's perspective - Prof. Peter Furness

A test may have a variety of different purposes, including making or excluding a diagnosis, guiding further investigation, evaluating prognosis, guiding treatment, monitoring treatment, population screening, keeping the doctor (or the patient) happy and avoiding litigation.

Because of this variety and complexity, evaluation of diagnostic tests is more complex than drug evaluation; a test may be effective for one purpose or in one population but not another. A multidisciplinary approach is required to examine all the issues surrounding a particular test. Although there are currently a number of national bodies involved in test evaluation, they concentrate on the thorough evaluation of a relatively small number of high-impact areas, leaving numerous 'smaller' questions unanswered. Their coverage is by no means exhaustive and each has its own specific remit and perspective, making it almost impossible to make an evidence-based evaluation or comparison between tests. There is also a lack of communication between the different bodies and it is unclear where ultimate responsibility for test evaluation lies.

This problem may in part be addressed by generating an authoritative list, such as a 'diagnostics formulary', of all laboratory investigations that should be available to all NHS patients, specifying under what clinical circumstances a test is used and listing complications, contra-indications, cross-reactions, supporting tests, sample requirements, expected confidence limits, interpretation and other parameters of the test. Such a database would require a system of horizon-scanning to keep it up to date and a process to identify missing evidence and to promote relevant research. A database such as this could be developed de novo or added to existing systems. Ultimately, submission of results to this database should be a pre-requisite for funding of any relevant research project, just as 'publication in peer-reviewed journals' is at present.

It is inevitable that such a process will encounter problems with a lack of evidence, especially evidence of clinical utility. Stagnation could result if the level of evidence demanded were to equal that required for a NICE evaluation. The rigour of the evaluation should therefore be proportionate to the questions – financial as well as therapeutic – being addressed, and the published outcome should include an assessment of the quality of the evidence. The 'gene dossier' system of the UK Genetic Testing Network provides an example of a relatively simple system of evaluation that has proved itself fit for purpose in the context in which it is used.

This database would be of value to a range of stakeholders including clinicians, pathologists, managers, commissioners, clinical auditors, researchers, industry, IT engineers and, of course, patients. However, several key questions remain: who will develop it and maintain it; who will pay for it; what standards should be used; and who will decide if it's a success?

2.2 Introduction: a public health perspective - Dr Ron Zimmern

Substantial consideration has already been given to the evaluation of genetic tests. The principles established in the context of genetic testing may be applied as a paradigm for test evaluation in general, as the problems and issues concerning the evaluation and regulation of genetic tests are applicable to all forms of diagnostics and biomarkers. Failure to address such matters will soon (if not now) be of major public health concern.

When discussing test evaluation, it is important to distinguish between an assay, which is a method to analyse or quantify a substance in a sample, and a test, which is a procedure that makes use of an assay in the context of:

- a particular disease
- in a particular population
- for a particular purpose

This distinction has practical implications. Whilst it may be relatively straightforward to make a technical evaluation of an assay, the evaluation of a test is more complex and inherently less susceptible to standardisation. Definition of the precise purpose (or purposes) of a test is crucial to evaluation as, like any other health care intervention, its effectiveness must be measured by the extent to which it achieves its intended objectives or purposes. A test may therefore be effective for one purpose, or in a particular population, but ineffective and of little value when used for a different purpose.

The ACCE framework, originally developed for evaluation of genetic tests, is applicable to all forms of molecular diagnostics and biomarkers. Broadly, it outlines four key factors for evaluation:

- Analytic validity of a test defines its ability to measure accurately and reliably the component of interest - its technical performance
- Clinical validity of a test defines its ability to detect or predict the presence or absence of clinical disease or predisposition to disease
- Clinical utility of a test refers to the likelihood that the test will lead to an improved outcome
- Ethical, legal and social implications (ELSI) of a test

It is particularly important to understand that clinical validity is more than just evidence of biomarker-disease association; it must also include the evaluation of clinical test performance (sensitivity, specificity, positive and negative predictive value and other test parameters) and its impact on health outcomes.

The Cooksey Report (2006)² identified two gaps in the translation of science into clinical benefit: first, “bench-to-bedside” translational research; and second, the “research-to-practice” translation of science and technology into clinical services. These have also been referred to respectively as Type 1 and Type 2 translation. At present, the bulk of research funding is directed at the first gap. In relation to diagnostic tests and biomarkers, there is growing interest in establishing relationships between biomarker and disease and in developing diagnostic products and services based on the science. By contrast there is a policy void in deciding how to undertake the evaluation and regulation of these diagnostic technologies for clinical use.

Identifying a biomarker-disease association is necessary, but not sufficient, for effective clinical performance. Currently, there are no systematic processes and platforms for generating clinical data (akin to Phase III pharmaceutical studies) to inform test evaluation, and no agreement about whose responsibility it should be to provide the resources for, or to carry out, such studies. Furthermore, there is no consensus nationally or internationally about the standards required and no organisations have a specific responsibility for systematically analysing and documenting results from studies of diagnostics and biomarkers.

Policy reform is urgently needed to establish systems and resources to generate evidence of test performance, and to agree the respective roles and responsibilities of government, statutory regulators, public bodies, academia and the commercial sector. Systems should be established to ensure that the data are appropriately analysed and evaluated against agreed standards and that the evidence is placed in the public domain. Funders and reimbursers of health services and clinicians should be discouraged from using tests that are not backed by appropriate clinical evidence. The role of statutory regulators should be confined to ensuring the safety of all tests and biomarkers, and that evidence in relation to test performance (or lack of it) is placed in the public domain.

2.3 Evaluating new laboratory investigations in the context of NHS commissioning - Prof. Chris Price

Commissioning in the Department of Health is driven by finding the best value for patients and taxpayers, “meaning (i) the best possible health outcomes, (ii) the best possible healthcare, (iii) within the resources made available by the tax payer” (Commissioning in the NHS, Department of Health, 2006). However, this type of evidence for laboratory diagnostic medicine is generally poor; incentives to develop such evidence are absent and typically reimbursement is based on cost not value. Positive health outcomes are primarily achieved through caring for individual patients whilst maximising benefit and minimising risk at a reasonable cost. Within this framework, diagnostic tests should be evaluated on the basis of how they expedite or optimise the care pathway (from presentation and diagnosis through to treatment and monitoring) and maximise the benefit of the test on the health outcome. Sackett and Haynes, in their paper entitled ‘The Architecture of Diagnostic Research’¹³, expressed this very simply with the question “Do patients undergoing the diagnostic test fare better than similar untested patients?”

Currently, there is a gap in translational research between the identification of a biomarker associated with a particular disease and proving that patients who are tested for this biomarker have better outcomes than those who are not; studies of laboratory diagnostic accuracy are not sufficient to justify clinical use. Evidence-based decisions require information regarding the diagnostic performance, clinical impact, organisational impact and cost effectiveness of a test as well as its technical performance.

The effectiveness with which a new test achieves a specified health outcome should be assessed by comparison with the current methodologies in the same patient population. A hierarchy exists to assess the strength of evidence associated with any test¹⁴, ranging from unsystematic clinical observations, through randomised trials and systematic reviews. Within any of these studies, various types of valid health outcomes may be used to assess the clinical, operational and economic effect of a test:

Type of outcome	Clinical	Operational	Economic
<i>Hard</i>	Morbidity, mortality, disability	Time to treatment	Cost /QALY, cost/diagnosis, cost/treatment
<i>Soft</i>	Patient satisfaction	Waiting time	–
<i>Surrogate</i>	Complication or readmission rate	Length of stay	Length of stay

Commissioners are interested in the delivery of high-quality care pathways; they are therefore interested in how diagnostic tests contribute to the pathway, and want to be assured that the diagnostic test is offering good value for money both in the delivery of the test, and in its application throughout the care pathway.

2.4 Developing Labtests Online - Dr Stephen Halloran

Labtests Online (www.labtestsonline.org.uk) is a public online database of laboratory diagnostic tests available in NHS laboratories, that is designed to help patients understand why particular tests might be requested and how their test results might help a doctor diagnose and treat their disease. In England, over 700 million laboratory tests are carried out each year, more than a third of which are generated in primary care. However, according to the 2005 Health Commission Report, a third of patients said that the results of diagnostic tests were not explained by their doctors in a way that they could understand.

Based on a similar American website, provided by the American Association of Clinical Chemistry since 2001, the UK version was set up and funded initially by a 3 year grant from the Health Foundation with current support from the RCPATH. Labtests Online UK launched in June 2004 and, as of January 2008, has received 4 million visits, of which the majority go directly to the site. Although the resource is currently primarily aimed at patients, more than 30% of site visitors are healthcare professionals.

Currently Labtests Online covers only established laboratory tests, categorised by test type, conditions/diseases and screening programmes. It includes information about the purpose, method and interpretation of each test, common questions and links to relevant sources of further information.

The text developed for Labtests Online UK is now being used in the development of similar sites across Europe. These sites are being supported by the European Diagnostic Manufacturers Association. Labtests Online is still developing, and consideration is being given to including a section with more detailed information suitable for healthcare professionals. It is possible that this platform could provide the infrastructure for additional databases for test evaluation, as advocated at this meeting by Professor Furness and others.

2.5 Evaluation of genetic tests: the experience of the UKGTN - Dr Mark Kroese

The UK Genetic Testing Network (www.ukgtn.nhs.uk) was established in 2002 to promote high quality, equitable services for patients and their families who require genetic advice, diagnosis and clinical management. Its core functions include evaluation of new genetic tests, approval of laboratories for membership and audit of the laboratory services provided. It is a voluntary NHS organisation and its membership comprises 31 NHS molecular genetic laboratories (including all 22 regional genetic laboratories) and one UK-based private company. The Network is managed by a number of co-ordinators and advised by a panel of scientific, clinical and public health experts. As part of the process of evaluation and approval, the UKGTN publish test criteria for specific genetic tests that are in the public domain. Laboratories wishing to provide a testing service to other members of the group must complete an application form for evaluation, detailing evidence associated with the test.

Currently, the UKGTN has limited itself to the evaluation of new genetic tests for single gene disorders where nucleic acids are the analyte. It has developed a pragmatic, multidisciplinary approach to evaluating the potential clinical effectiveness of each test using a process called the 'Gene Dossier', which is based on the ACCE framework. In the period 2004-2007, 89 gene dossiers were submitted of which 70% were accepted. Of those that were rejected, the most common reasons for rejection was insufficient evidence regarding clinical validity and clinical utility. These applications often have insufficient evidence of the advantages of the test compared to the conventional pathway. In general, applications did not include a thorough economic assessment or sufficient consideration of the ethical, legal and social implications of the test.

In 2005 the UKGTN modified the gene dossier to reflect important features of evaluation such as target population, disease and purpose of testing and has since reviewed and approved 49 newly developed genetic tests. Recently, the UKGTN has widened the scope of its activities by inviting cytogenetic laboratories to join the network, and is prepared to offer an advisory role to help establish a process for evaluation of genomic-based tests and complex biomarkers used in the NHS.

The experience of the UKGTN has highlighted a number of key areas which must be addressed for the wider evaluation of laboratory diagnostics tests:

- Disorder, healthcare setting and clinical context need to be clearly defined and fully understood as they are crucial to the evaluation process; evaluation of a test outside a care pathway is essentially meaningless
- Establishing clinical validity data for tests is challenging but necessary; clinical utility – whilst ultimately a subjective, context-dependent judgement – is the most important domain in the final decision making process
- Systems and platforms for the generation of such data are needed
- A multidisciplinary approach to evaluation is required

2.6 What information does the health service need on diagnostics, and how should it be delivered? (Blood in; information out)

- Sir Muir Gray

Knowledge of the context and clinical pathway in which a test will be used is crucial to evaluation. Understanding the purpose of a test allows the benefits to be balanced against the hazards in order to decide an appropriate level of investment. It is recognised that increasing investment in a particular intervention is accompanied initially by increased benefits. However, the amount of benefit per unit of investment slows and eventually levels off as investment increases, whilst the harms continue to increase linearly with investment. For every intervention, therefore, there is a point at which investment of resources achieves the optimum balance between benefits and harms.

Before deciding how much to invest in a test, it is crucial to understand its clinical utility, specifically:

1. Are the interventions being offered all likely to confer a high value for this group of patients? Put another way, is there a good balance between benefit and harm at an affordable price?
2. Are the patients who are most likely to benefit (and least likely to be harmed) from the high value interventions clearly defined?

Although clinical utility can be assessed in a well-defined group of patients, in practice, it may be difficult to limit a particular intervention to that group – often a procedure is introduced for a limited group of patients but, over time, its use drifts into other groups.

As tests become more numerous and complicated, clinical judgement is no longer sufficient for evaluating which test is appropriate within each clinical context. Therefore expert guidance is needed, to integrate and coordinate information across disciplines. A number of online knowledge databases have been developed to deal with the challenges of 21st century medicine and a changing workforce, including the Map of Medicine (www.mapofmedicine.com), which details different patient pathways, the National Library for Health (www.library.nhs.uk) and the NHS National Knowledge Service (www.nks.nhs.uk). Healthcare systems are best managed by networks such as these, rather than by bureaucratic organisation. An example of an effective network is provided by the National Down's Syndrome Screening Programme, which is a knowledge-based organisation that ensures consistent service across multiple bureaucracies with shared aims.

2.7 Additional short presentations

The Summit was further assisted by three short reports from delegates involved in existing projects relevant to the evaluation of laboratory diagnostic tests:

Ms Diana Garnham, The Science Council

The Science Council, a membership organisation representing learned societies and professional institutions across the breadth of science in the UK, recognises the great importance of properly evaluating diagnostic tests. It has recently completed a report entitled 'Integration and Implementation of Diagnostic Technologies in Healthcare', in which it recommends that all available evidence about performance and utility of a new diagnostic test be systematically assessed. The importance of an appropriate structure

for the assessment and implementation of new diagnostic technologies in the NHS was highlighted, as was the need for a skilled workforce to deliver reliable and consistent diagnostic services.

Prof. David Melzer, Peninsula Medical School

Results were briefly outlined from a recently completed Wellcome Trust funded project to assess how genetic tests for common disease susceptibility enter routine clinical practice. The project considered the need for appropriate evaluation before such screening tests either enter routine clinical practice or are advertised directly to the public. Issues relating to the regulation of these tests include the international nature of direct-to-consumer testing as well as gaps in existing legislation and policy. Amongst both clinicians and patients, there was near universal support for the need to improve the generation of good clinical evidence on tests, and for this evidence to be made easily and quickly available to doctors, patients and consumers.

Dr Rick Jones, Yorkshire Centre for Health Informatics

Numerous ongoing NHS, Connecting for Health and Department of Health sponsored projects to create information technology resources for pathology were outlined, including the National Pathology Catalogue, Harmony, Pathology Subsets Group, Clinical Best Practice, National Laboratory Handbook, Primary Care Benchmarking and Lab to Lab. The need for integration and coordination of these and other databases via a central hub was noted. Effort will be needed to define a suitable and supportable informatics architecture which takes account of global coding and classification standards (SNOMED, HL7) to ensure interoperability between clinical systems and knowledge sources, including the proposed evidence base for diagnostic tests.

3 Discussion

The discussion section summarises the views of delegates grouped into categories for easy reference. Individuals have not been identified in this section.

3.1 Database of tests

It was unanimously agreed by all delegates that there was an urgent need for a single, publically available database aimed at professionals, containing evidence for the validity and utility of laboratory tests. A single system would provide integration and co-ordination across the different subspecialties within pathology.

This database should have a network structure such that it links multiple sources of information together. Initially, existing databases (specifically Labtests Online and the National Pathology Catalogue currently under development at the Department of Health) should be considered as the infrastructure on which to build such a function.

Evidence for a particular test should include:

- biomarker-disease association
- technical performance (analytical validity)
- assessment of clinical validity and test purpose
- assessment of clinical utility and impact
- use of the test in different clinical situations (tied to care pathways) and for particular purposes
- any specific circumstances where use of the test has been reviewed and is regarded as unjustified

Within this database, knowledge gaps should be explicitly highlighted, so as to make it clear where evidence is lacking and more data are required. The evidence should be clearly set out and the purpose or objective of the test on which the evidence is based should be specifically noted, so an overall recommendation for its use (akin to the categorisation used in the British Medical Journal Clinical Evidence) can be made.

A central hub would be needed to steer the workings of the database. This hub should be professionally managed and led, and would critically examine evidence for the use of particular tests. It would need both capital and recurrent revenue resources. Broadly, two options were suggested:

1. To establish a new national body, similar to the UKGTN, with a specific remit to collate evidence associated with laboratory diagnostics tests
2. To place the responsibility with professional societies, such as the RCPATH, the Association of Clinical Biochemistry or the Joint Royal Colleges

Delegates noted with interest the various initiatives being established by the Department of Health to collate laboratory test data, but they were uncertain whether, as presently conceived, these systems would be capable of delivering a database of sufficient scale and complexity. They felt that a scoping document outlining the purpose and parameters required of the database would need to be produced, and that it was probably best that the development and implementation of the database be put out to tender.

Delegates were unable to decide whether tests sold directly to the patient ('direct-to-consumer') should be included on the database. There was a general feeling that it should be limited to tests available through healthcare practitioners specifically for medical purposes, thus excluding so-called 'lifestyle' tests. It was agreed that this matter needed further consideration once the database had been established. Notwithstanding these comments, there was support for the view that there were specific issues in relation to the use of tests supplied directly to the public, which would need further special consideration.

3.2 Generation of data for test evaluation

Many delegates noted that, where evidence was missing for the clinical validity and utility of tests, it was currently unclear whose responsibility it was (or should be) to generate new data. Whereas significant research funding was available to develop biomarker-disease associations, and the Health Technology Assessment programme allowed for research proposals to analyse clinical practice and public health outcomes associated with a particular test, there was a gap in the provision of a system for the systematic evaluation of diagnostics and biomarkers.

Since these data, which are analogous to Phase III clinical trials, are generally expensive and slow to generate, there was a general (but not unanimous) view that public-private partnerships between industry and clinical scientists would be the best way to address this problem. A way was needed to bring industry and the public sector together to discuss such issues and to determine the roles and responsibilities of the various stakeholders. Some delegates voiced the opinion that, just as with pharmaceuticals, the responsibility for generating the evidence for validity and utility should fall primarily upon the test developer. Others expressed concern that such a requirement would inhibit the evaluation of new tests, especially those that lack a commercial sponsor with a large financial interest. In this respect it was noted that there was little incentive for the commercial sector to generate the required evidence, firstly because it was not required in other health systems, and secondly because the NHS focussed on cost rather than value (Carter Report 2006¹⁵).

In the longer term, the issues of intellectual property and reimbursement surrounding diagnostics would need to be addressed, so that diagnostic companies would have an adequate incentive both to develop and thoroughly evaluate new diagnostic tests.

The need for horizon scanning was discussed within the context of a database of laboratory diagnostic tests. It was agreed that there were existing organisations that attempted to perform this function, but that their work would benefit from greater co-ordination and explicit engagement with professional organisations. Any developments identified as a result of horizon scanning should be communicated to those with oversight of the database and ultimately incorporated into it. Steps had to be taken to ensure that the horizon scanning of diagnostics and biomarkers should be explicitly established as part of their brief. It was also necessary then to feed the results to those responsible for overseeing the diagnostics database.

Delegates conceptualised the proposed database to be like the current database of active clinical trials, whereby no trial data could be published unless it was entered into the database. It was envisaged that test developers would register all new tests, and that test purchasers should only consider tests supported by sufficient evidentiary data.

3.3 Test evaluation

Much of the discussion focused on the importance of understanding and clearly defining the purpose(s) of a test in terms of its usage within a clinical pathway and the patient population for whom it was intended. Delegates agreed that, whether conceptualised as the need to define 'purpose' more explicitly or to be clear about the 'clinical pathway' concerned, these ideas were more or less equivalent and were central to evaluating the clinical validity and utility of a test.

The enormity of the task of evaluating and regulating all diagnostic tests necessarily required some form of prioritisation process. Although it was the test that had to be the subject of the evaluation rather than the disorder, it was important to evaluate the test in the context of the disorder and the clinical pathway. This insight might also inform how tests could be prioritised for initial evaluation. Those tests that were part of the NHS 'Big 50' - a list of the 50 most important clinical conditions currently encountered in the NHS - might be the subject of the initial batch for aggregation of evidence for the suggested database.

In an ideal world tests should only be available for routine clinical use following the provision of adequate evidence of clinical validity and utility. A number of bodies were already involved in test evaluation [the National Institute for Health Research Health Technology Assessment (HTA) programme, the National Institute for Clinical Excellence (NICE), NHS PASA Centre for Evidence-based Purchasing (CEP) and independent research programmes by the MRC, universities and others] but none had a specific brief for diagnostics. The number of evaluations carried out by these bodies was small, and evidence from the UKGTN showed that in many cases it was the lack of data on which to carry out the evaluation, as distinct from having a body able to carry out the evaluation, that was the limiting factor.

Although the extremely thorough methods used by HTA and NICE were considered to be a gold standard, a faster, more pragmatic approach was required for laboratory test evaluation due to the number of tests involved. A model for this might be the exemplary method used by the UKGTN for the evaluation of genetic tests. A number of options were presented through which new evaluations could be carried out:

1. NICE assessments were agreed to be the gold standard, but were very slow. Their procedures were rigorous and therefore time consuming. Whilst specific tests of major national importance could be dealt with by this system, it was not felt that NICE would be able to deal with the volume that diagnostic test evaluation required. Nevertheless delegates agreed that it had a specific place within a general system for test evaluation.
2. NICE or the HTA programme could establish under their aegis a specific diagnostic assessment process akin to that which now exists for interventional procedures guidance.
3. A new body, akin to the UKGTN, could be set up specifically to evaluate the evidence associated with all laboratory diagnostic tests.
4. The NHS PASA Centre for Evidence-based Purchasing (CEP) currently carries out fast technology evaluations and could potentially be augmented to give it a wider remit.

Delegates felt that all these options should be put forward for future discussion.

3.4 Regulatory Framework

The role of the statutory regulators in laboratory diagnostic test evaluation was discussed. It was agreed that transparency was a pre-requisite for test evaluation. Some delegates expressed the view that it was necessary to require all evidence of test validity and utility to be placed in the public domain. It was noted that the European Directive on In-vitro Diagnostic Medical Devices (EC98/79/EC) contained an express obligation upon all parties to observe confidentiality, which might preclude transparency from those marketing diagnostic tests. These obligations had been incorporated into UK law by the Medical Devices Regulations 2002. To the extent that any conflict existed, it was suggested that the confidentiality clause (Article 19) of the European In-Vitro Diagnostics Directive would need to be modified to allow regulation requiring that evidence relating to test performance (or lack of it) be placed in the public domain. Given the global nature of the testing market, it was important to have European recognition of the UK regulatory process, and the UK Government and relevant statutory authorities should therefore be urged to make submissions in these terms to the appropriate European authorities.

Collecting evidence on the clinical performance of a test was hampered by a lack of funding for such studies, and by the absence of a regulatory requirement for companies to undertake them. The lack of incentives for industry to generate key clinical data was thought to be one of the reasons for the lack of evidence of clinical validity and utility. Current regulations made it sufficient for a test provider only to show evidence of laboratory test performance (analytical validity). They were only obliged to go further if they made a clinical claim, but even then it was unclear what sort of evidence was required by the regulators. It was noted that commercial organisations complained that, unlike pharmaceutical products, it was unclear what evidence they needed to provide, and to whom, in order to facilitate the use of a new test within the NHS. It was possible that an association between a biomarker and a particular disease was generally taken as sufficient for a test, rather than evidence of good test clinical performance. If details of the clinical efficacy of a test were provided as a pre-requisite for CE-marking (together with the removal of the obligation for confidentiality) this would increase transparency and may go some way to addressing this problem, but this seemed unlikely to be achieved in the short term. Reform of the intellectual property system to provide appropriate incentives might also be needed.

Statutory regulators had an important role to ensure the safety of all tests and biomarkers. Current regulations did not stratify the evidential burden in relation to pre-market authorisation by risk, except in the crudest way. Delegates agreed that a more responsive and realistic categorisation of risk assessment during pre-market approval was needed to protect the public from serious harm. There was no call from delegates to require positive evidence of clinical validity and effectiveness before allowing a test onto the marketplace, but the fact that such evidence did not exist should be made available for all to see. The role of the statutory regulator was not to prevent tests with little evidence from being sold; it was rather to ensure that whatever evidence was available should be put in the public domain. The regulators also had a responsibility to risk stratify diagnostic products so

that only unsafe (as distinct from ineffective) diagnostic products were precluded from sale through pre-market assessment. Delegates also discussed whether such pre-market risk stratification should be carried out by an independent scientific committee.

The role of the statutory authorities in regulating over-the-counter tests sold directly to the public was not discussed in any detail, but there was acknowledgement that these posed different types of risks and issues, all of which would need more detailed consideration.

4 Recommendations

1. A new body should be established to ensure the evaluation of laboratory diagnostic tests and the creation of a database of new and existing laboratory tests.
2. This body might be established de novo along the lines of the UK Genetic Testing Network, or the responsibility could be placed with existing professional societies such as the Royal College of Pathologists, the Association of Clinical Biochemistry or the Academy of Royal Colleges.
3. A publically available database of existing and new diagnostic laboratory tests should be set up containing evidence, or explicitly the lack of it, for the validity and utility of clinical laboratory tests.
4. Where a test evaluation had already been carried out and published by an appropriate agency, it should be linked to the database.
5. Where evidence is missing for existing tests, particularly evidence of clinical validity and utility, consideration should be given to funding the necessary studies.
6. Policy makers and all stakeholders should be encouraged to address issues around funding and gathering the necessary evidence for the clinical evaluation of new and complex biomarkers, and should consider the establishment of private-public partnerships to increase industry involvement.
7. An independent expert body should be responsible for the evaluation of the evidence for test performance and making recommendations about clinical use.
8. Commissioners and health care professionals should be encouraged to use only those tests where sufficient evidence of clinical performance exists.
9. Statutory regulators should be empowered to require that evidence (or lack of) relating to test performance be placed in the public domain.
10. A more responsive and proportionate risk assessment during pre-market approval is needed to ensure patient safety.

5 Appendices

Appendix 1 Glossary of terms

ACCE: a framework to evaluate the clinical value of a genetic test that includes Analytical validity, Clinical validity, Clinical utility and ELSI.

CE mark: a mandatory conformity mark on many European products in the European Economic Area. CE marking indicates conformity with essential health and safety requirements set out in European Directives and requires self-certified documented proof that the item meets the relevant requirements.

CEP: the Centre for Evidence-based Purchasing, works to ensure that the NHS in England gets the best value for money when purchasing goods and services.

ELSI: Ethical, Legal and Social Implications.

HL7: Health Level Seven, a not-for-profit organization involved in development of international healthcare information exchange standards. The standards, which support clinical practice and the management, delivery, and evaluation of health services are a core component of the NHS Information Strategy.

HTA: Health Technology Assessment, a multidisciplinary process that produces studies the effectiveness, costs, and broader impact of health technologies.

MRC: Medical Research Council, a national organisation funded by the UK government that promotes research into all areas of medical and related science

NHS: National Health Service, the publicly funded health care system in England.

NICE: National Institute for Clinical Excellence, an independent organisation responsible for providing national guidance on promoting good health and preventing and treating ill health within the NHS.

PASA: Purchasing and Supplies Agency, executive agency of the Department of Health that advises the NHS on the most effective use of its resources by getting the best possible value for money when purchasing goods and services.

Phase I, II and III clinical trials: different stages of human trials that all therapeutics must pass in order to be licensed for use.

- Phase I trials assess safety and drug dose in a small number of people.
- Phase II trials assess efficacy in a small number of target patients.
- Phase III trials compare effectiveness and safety with the existing 'gold standard' in a large number of patients.

QALY: Quality-Adjusted Life Years, a way of quantifying the benefit of a medical intervention in terms of both the quality of life and the length of survival.

RCPATH: Royal College of Pathologists, a professional membership organisation with charitable status that is concerned with ensuring the quality of all matters relating to the science and practice of pathology.

SNOMED: the Systematised Nomenclature of Medicine, a clinical terminology that will be used by all computers in the NHS to facilitate communications between healthcare professionals in clear and unambiguous terms.

Type 1 translation: basic research needed to advance experimental biomedical studies into preclinical testing in humans.

Type 2 translation: focuses on strategies to ensure that new clinical research findings are implemented appropriately.

UKGTN: United Kingdom Genetic Testing Network, a collaborative network of UK-based laboratories, clinicians, commissioners and patient representatives that aims to provide equal access to high quality genetic testing services for patients with single-gene, germline disorders across the whole of the UK.

Appendix 2 Summit Programme

DAY 1 – Monday 14 January		
1600 - 1630	Registration	
1630 - 1635	Welcome	Ron Zimmern
1635 - 1715	Introduction of delegates	Ron Zimmern
1715 - 1745	Introduction: the pathologist's perspective	Peter Furness
1745 - 1815	Introduction: the public health perspective	Ron Zimmern
1815 - 1900	General discussion	
1900 - 1930	Break	
1930	Social dinner	
DAY 2 – Tuesday 15 January		
0800 - 0900	Breakfast	
0900 - 0930	Evaluating new laboratory investigations in the context of NHS commissioning	Chris Price
0930 - 1000	Developing Labtests Online	Stephen Halloran
1000 - 1030	Break	
1030 - 1100	Evaluation of genetic tests: the experience of the UKGTN	Mark Kroese
1100 - 1130	What information does the health service need on diagnostics, and how should it be delivered?	Muir Gray
1130 - 1230	General discussion	
1230 - 1330	Lunch	
1330 - 1515	Discussion session: a) The need for a database of approved laboratory investigations b) Horizon scanning for new tests c) Generating the data - whose responsibility? d) Evaluating test validity and utility - whose responsibility?	Peter Furness & Ron Zimmern
1515 - 1545	Final wrap-up - Agreeing recommendations	Peter Furness & Ron Zimmern

Appendix 3 Summit Participants

(a) Delegates

NAME (* Speaker)	TITLE
Dr Ian Barnes	Head of Pathology, Clinical Biochemistry, Leeds General infirmary
Dr Marie-Ange Baucher	Junior Consultant, Biotechnology Division, Directorate for Science, Technology & Industry, Paris, France
Mrs Valerie Bevan	Director, Evaluations & Standards Laboratory, Centre for Infections, London
Dr David Brown	Consultant Medical Virologist, Virus Reference Department, Health Protection Agency, London
Prof. Martin Buxton	Director, Health Economics Research Group, Brunel University
Dr John Crolla	Chairman - Joint Committee on Medical Genetics, Wessex Regional Genetics Laboratory, Salisbury District Hospital
Prof. Peter Farndon	Chairman, UKGTN, Birmingham Women's Hospital, Birmingham
* Prof. Peter Furness	Vice-President, Royal College of Pathology, London
Mrs Katie Garner	Pathology Cluster Manager, NHS PASA Centre for Evidence-based Purchasing (OEP), London
Ms Diana Garnham	Chief Executive, The Science Council, London
Ms Carol George	Head of Policy, PHG Foundation, Cambridge
* Sir Muir Gray	Director, National Knowledge Service, & Chief Knowledge Officer to the NHS, Oxford, NHS National Knowledge Service, Oxford
* Dr Stephen Halloran	Consultant Clinical Biochemist, GMEC/BCSP, Post-Graduate Medical School, University of Surrey, Guildford
Prof. Steve Humphries	BHF Prof. of Cardiovascular Genetics, Centre for Cardiovascular Genetics, BHF Laboratories, London
Dr Rick Jones	Associate Clinical Director, Yorkshire Centre for Health Informatics, University of Leeds
Prof. Noor Kalsheker	Prof. of Clinical Chemistry, University Hospital, Nottingham
* Dr Mark Kroese	Public Health Advisor - UK Genetic Testing Network, Peterborough Primary Care Trust
Mr Stephen Lee	Principal Medical Device Specialist, Biosciences & Implants Unit, Medicines & Healthcare Projects Regulatory Agency, London
Dr Susanne Ludgate	Clinical Director - Devices, Medicines & Healthcare Projects Regulatory Agency, London
Dr Anne Mackie	Director, Screening Programmes, UK National Screening Committee, London
Prof. David Melzer	Prof. of Epidemiology & Public Health, Peninsula Medical School, Exeter
Dr Claire Packer	Director, National Horizon Scanning Centre, Department of Public Health, Birmingham University
Ms Margaret Parton	CEO, NHS National Technology Adoption Hub, Manchester Royal Infirmary, Manchester
Dr Christine Patch	Consultant Genetic Counsellor/Manager, Department of Clinical Genetics, Guys Hospital, London
Prof. Munir Pirmohamed	NHS Chair of Pharmacogenetics, Department of Pharmacology & Therapeutics, Liverpool University
Dr Aarathi Prasad	Public Liaison, Sense About Science, Shaftesbury Avenue, London
* Prof. Christopher Price	Visiting Prof. in Clinical Biochemistry, Oxford University
Prof. Julian Sampson	Prof. & Honorary Consultant in Medical Genetics, Institute of Medical Genetics, University of Wales College of Medicine, Cardiff, Wales

Dr Nigel Sansom	Senior Manager for Technology Introduction, National Innovation Centre, NHS Institute for Innovation & Improvement
Dr Nicholas Summerton	Consultant Clinical & Public Health Advisor, NICE, London
Dr Jenny Taylor	Programme Director, Genomics & Pathology Theme, Oxford Biomedical Research Centre, The Wellcome Trust Centre for Human Genetics, Oxford
Prof. Tom Walley	Director of HTA Programme, Old Infirmary, Liverpool
Dr Virginia Warren	Public Health Physician, BUPA, BUPA House, London
Dr Ian Watson	Consultant Biochemist, Department of Clinical Biochemistry, University Hospital Aintree, Liverpool
* Dr Ron Zimmern	Executive Director, PHG Foundation, Cambridge

(b) Supporting Team

Dr Maria Adams	Project Coordinator, PHG Foundation, Cambridge
Ms Alison Hall	Project Manager (Law), PHG Foundation, Cambridge
Dr Caroline Wright	Project Manager (Science), PHG Foundation, Cambridge

Appendix 4 Speaker Biographies

Prof. Peter Furness

Peter Furness is a Consultant Histopathologist and Honorary Professor of Renal Pathology in Leicester. He has a longstanding interest in laboratory quality. In 2002 he was elected to Council of the Royal College of Pathologists and in 2005 he was elected Vice President. During this period one of his activities has been to initiate and lead an RCPATH project on how new laboratory investigations might be better evaluated for NHS use.

Sir J A Muir Gray

Muir Gray is Director of the National Knowledge Service, Chief Knowledge Officer for the NHS and is closely involved in the provision of knowledge, not only to clinicians, but also to patients and those who manage healthcare. He is the author of Evidence-Based Healthcare, of which the third edition is in preparation, and joint author of The Oxford Handbook of Public Health Practice.

Dr Stephen Halloran

Stephen Halloran is a consultant biochemist at the Royal Surrey County Hospital and a Senior Research Fellow at the University of Surrey. He is director of GMEC (Guildford Medical devices Evaluation Centre) which provides an evaluation service for the DH Centre for Evidence-based Purchasing. He is also Managing Editor of Labtests Online UK, a web-based laboratory test information service for patients.

Dr Mark Kroese

Mark Kroese is a consultant in public health medicine at Peterborough Primary Care Trust, who is seconded part-time to the PHG Foundation to work on the evaluation and regulation of genetic tests and biomarkers. He has been involved in the development of genetic test evaluation and its implementation for several years, and has practical experience of performing genetic test evaluations. In March 2006, he was appointed Public Health Advisor to the UK Genetic Testing Network.

Prof. Chris Price

Chris Price has been a Visiting Professor in Clinical Biochemistry at the University of Oxford since 2002 and a member of the Carter Review team. Between 2002 and 2005 he was Vice President of Outcomes Research for the Diagnostics Division of Bayer Healthcare, and prior to that he was Director of Pathology at Barts and The London NHS Trust. His main professional interests lie in the area of evidence-based laboratory medicine and outcomes research. He is one of the editors of Evidence-Based Laboratory Medicine, the second edition of which was published in 2007.

Dr Ron Zimmern

Ron Zimmern is Director of the newly established charitable Foundation for Genomics and Population Health, the successor to the Public Health Genetics Unit which he established in Cambridge in June 1997. He is also an Associate Lecturer at the University of Cambridge and an Honorary Consultant in Public Health Medicine at Addenbrooke's Hospital. He was Chairman of the Diagnostic and Screening Panel of the UK Health Technology Assessment programme until end 2007, serves on the Genetics Commissioning Advisory Group and the Steering Group for the National Genetic Testing Network at the Department of Health, and was on the Council for the British Society of Human Genetics until March 2007.

Appendix 5 List of Papers Tabled

- ***Evidence and evaluation: building public trust in genetic tests for common diseases***, DRAFT Research Project Report, Melzer, Hogarth, Liddell, Ling, Sanderson & Zimmern, January 2008
- ***Integration and implementation of diagnostic technologies in healthcare***, a report from the Science Council's Science in Health Group, The Science Council, January 2008
- ***Moving beyond ACCE: an expanded framework for genetic test evaluation***, a paper for the UK Genetic Testing Network, Burke & Zimmern, September 2007
- ***The evaluation of clinical validity and clinical utility of genetic tests***, Summary of an expert workshop, Kroese, Elles & Zimmern, PHG Foundation, September 2007
- ***The evaluation of novel genetic tests and complex molecular biomarkers***, A paper for the UK Genetic Testing Network, Zimmern & Kroese, PHG Foundation, July 2006

6 References

1. Royal College of Pathologists. Evaluating and introducing new diagnostic tests: the need for a national strategy. 2006.
2. Cooksey D. A Review of UK Health Research Funding. 2006.
3. Khoury MJ, Gwinn M, Yoon PW, Dowling N, Moore CA, Bradley L. The continuum of translation research in genomic medicine: how can we accelerate the appropriate integration of human genome discoveries into health care and disease prevention? *Genetics in Medicine* 2007;**9**:665-74.
4. Price CP. Evidence-based laboratory medicine: supporting decision-making. *Clin Chem* 2000;**46**:1041-50.
5. Burke W, Zimmern RL. Ensuring the appropriate use of genetic tests. *Nat Rev Genet* 2004;**5**:955-9.
6. Douglass CW. Evaluating diagnostic tests. *Adv Dent Res* 1993;**7**:66-9.
7. Kroese M, Zimmern RL, Sanderson S. Genetic tests and their evaluation: Can we answer the key questions? *Genetics in Medicine* 2004;**6**:475-80.
8. Haddow J, Palomaki G. ACCE: A Model Process for Evaluating Data on Emerging Genetic Tests. In Khoury MJ, Little J, Burke W, eds. *Human Genome Epidemiology*, Oxford University Press, 2004.
9. Gudgeon JM, McClain MR, Palomaki GE, Williams MS. Rapid ACCE: Experience with a rapid and structured approach for evaluating gene-based testing. *Genetics in Medicine* 2007;**9**:473-8.
10. Kroese M, Zimmern RL, Farndon P, Stewart F, Whittaker J. How can genetic tests be evaluated for clinical use? Experience of the UK Genetic Testing Network. *Eur J Hum Genet* 2007.
11. Burke W, Zimmern RL, Kroese M. Defining purpose: a key step in genetic test evaluation. *Genetics in Medicine* 2008;**9**:675-81.
12. Zimmern RL, Kroese M. The evaluation of genetic tests. *J Public Health* 2007;**29**:246-50.
13. Sackett DL, Haynes RB. Evidence base of clinical diagnosis: The architecture of diagnostic research. *BMJ* 2002;**324**:539-41.
14. Guyatt GH, Haynes RB, Jaeschke RZ, Cook DJ, Green L, Naylor CD et al. Users' Guides to the Medical Literature: XXV. Evidence-based medicine: principles for applying the Users' Guides to patient care. *JAMA* 2000;**284**:1290-6.
15. Lord Carter of Coles. Report of the Review of Pathology Services in England: an Independent Review for the Department of Health. 2006.