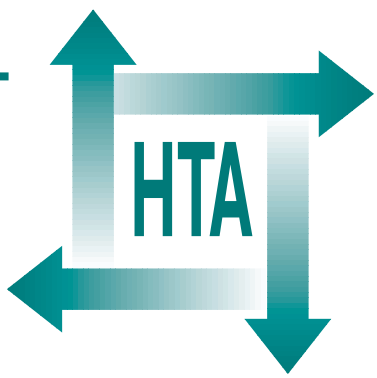
Health Technology Assessment 2003; Vol. 7: No. 10

Evaluation of molecular tests for prenatal diagnosis of chromosome abnormalities

GM Grimshaw
A Szczepura
M Hultén
F MacDonald
NC Nevin
F Sutton
S Dhanjal

Health Technology Assessment
NHS R&D HTA Programme



How to obtain copies of this and other HTA Programme reports.
An electronic version of this publication, in Adobe Acrobat format, is available for downloading free of charge for personal use from the HTA website (http://www.hta.ac.uk). A fully searchable CD-ROM is also available (see below).

Printed copies of HTA monographs cost £20 each (post and packing free in the UK) to both public **and** private sector purchasers from our Despatch Agents.

Non-UK purchasers will have to pay a small fee for post and packing. For European countries the cost is £2 per monograph and for the rest of the world £3 per monograph.

You can order HTA monographs from our Despatch Agents:

– fax (with **credit card** or **official purchase order**)
– post (with **credit card** or **official purchase order** or **cheque**)
– phone during office hours (**credit card** only).

Additionally the HTA website allows you **either** to pay securely by credit card **or** to print out your order and then post or fax it.

Contact details are as follows:

HTA Despatch
c/o Direct Mail Works Ltd
4 Oakwood Business Centre
Downley, HAVANT PO9 2NP, UK

Email: orders@hta.ac.uk
Tel: 02392 492 000
Fax: 02392 478 555
Fax from outside the UK: +44 2392 478 555

NHS libraries can subscribe free of charge. Public libraries can subscribe at a very reduced cost of £100 for each volume (normally comprising 30–40 titles). The commercial subscription rate is £300 per volume. Please see our website for details. Subscriptions can only be purchased for the current or forthcoming volume.

Payment methods

Paying by cheque
If you pay by cheque, the cheque must be in **pounds sterling**, made payable to *Direct Mail Works Ltd* and drawn on a bank with a UK address.

Paying by credit card
The following cards are accepted by phone, fax, post or via the website ordering pages: Delta, Eurocard, Mastercard, Solo, Switch and Visa. We advise against sending credit card details in a plain email.

Paying by official purchase order
You can post or fax these, but they must be from public bodies (i.e. NHS or universities) within the UK. We cannot at present accept purchase orders from commercial companies or from outside the UK.

How do I get a copy of *HTA on CD*?

Please use the form on the HTA website (www.hta.ac.uk/htacd.htm). Or contact Direct Mail Works (see contact details above) by email, post, fax or phone. *HTA on CD* is currently free of charge worldwide.

The website also provides information about the HTA Programme and lists the membership of the various committees.

Evaluation of molecular tests for prenatal diagnosis of chromosome abnormalities

GM Grimshaw[1*]
A Szczepura[1]
M Hultén[2]
F MacDonald[3]
NC Nevin[4]
F Sutton[5]
S Dhanjal[2]

[1] Centre for Health Services Studies, University of Warwick, Coventry, UK
[2] Biological Sciences, University of Warwick, Coventry, UK
[3] West Midlands Regional Genetic Laboratory and Consultancy Services, Birmingham Womens Health Care Trust, Birmingham, UK
[4] University of Belfast, Belfast, Northern Ireland
[5] Statistician, Learning and Skills Council, Coventry, UK

[*] Corresponding author

Declared competing interests of the authors: none

Published May 2003

This report should be referenced as follows:

Grimshaw GM, Szczepura A, Hultén M, MacDonald F, Nevin NC, Sutton F, *et al*. Evaluation of molecular tests for prenatal diagnosis of chromosome abnormalities. *Health Technol Assess* 2003;**7**(10).

Health Technology Assessment is indexed in *Index Medicus*/MEDLINE and *Excerpta Medica*/EMBASE. Copies of the Executive Summaries are available from the NCCHTA website (see opposite).

NHS R&D HTA Programme

The NHS R&D Health Technology Assessment (HTA) Programme was set up in 1993 to ensure that high-quality research information on the costs, effectiveness and broader impact of health technologies is produced in the most efficient way for those who use, manage and provide care in the NHS.

Initially, six HTA panels (pharmaceuticals, acute sector, primary and community care, diagnostics and imaging, population screening, methodology) helped to set the research priorities for the HTA Programme. However, during the past few years there have been a number of changes in and around NHS R&D, such as the establishment of the National Institute for Clinical Excellence (NICE) and the creation of three new research programmes: Service Delivery and Organisation (SDO); New and Emerging Applications of Technology (NEAT); and the Methodology Programme.

This has meant that the HTA panels can now focus more explicitly on health technologies ('health technologies' are broadly defined to include all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care) rather than settings of care. Therefore the panel structure has been redefined and replaced by three new panels: Pharmaceuticals; Therapeutic Procedures (including devices and operations); and Diagnostic Technologies and Screening.

The HTA Programme continues to commission both primary and secondary research. The HTA Commissioning Board, supported by the National Coordinating Centre for Health Technology Assessment (NCCHTA), will consider and advise the Programme Director on the best research projects to pursue in order to address the research priorities identified by the three HTA panels.

The research reported in this monograph was funded as project number 94/43/04.

The views expressed in this publication are those of the authors and not necessarily those of the HTA Programme or the Department of Health. The editors wish to emphasise that funding and publication of this research by the NHS should not be taken as implicit support for any recommendations made by the authors.

Criteria for inclusion in the HTA monograph series
Reports are published in the HTA monograph series if (1) they have resulted from work commissioned for the HTA Programme, and (2) they are of a sufficiently high scientific quality as assessed by the referees and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search, appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

The editors and publisher have tried to ensure the accuracy of this report but do not accept liability for damages or losses arising from material published in this report. They would like to thank the referees for their constructive comments on the draft document.

ISSN 1366-5278

Published by Core Research, Alton, on behalf of the NCCHTA.
Printed on acid-free paper in the UK by The Basingstoke Press, Basingstoke.

G

Contents

List of abbreviations

CVS	chorionic villus sampling
FISH	fluorescence *in situ* hybridisation
HTA	Health Technology Assessment
MLA	Medical Laboratory Assistant
MTO	Medical Technical Officer
NS	not significant
QALY	quality-adjusted life-year
QoL	quality of life
Q-PCR	quantitative polymerase chain reaction
ROC	receiver operating characteristic
STR	small tandem repeat
WTP	willingness to pay
VAS	visual analogue scale

All abbreviations that have been used in this report are listed here unless the abbreviation is well known (e.g. NHS), or it has been used only once, or it is a non-standard abbreviation used only in figures/tables/appendices in which case the abbreviation is defined in the figure legend or at the end of the table.

Executive summary

Background

Many women undergo prenatal tests for chromosome abnormalities in their baby, usually following identification of an increased risk of the baby having Down syndrome. One test that can show up abnormalities likely to lead to mental or physical handicap is done by sampling the amniotic fluid that surrounds the baby, usually around 14 weeks of pregnancy. There is a very small risk of miscarriage with this procedure and parents are warned of this. If the test shows there is chromosome abnormality, parents may want to discuss whether to continue with the pregnancy.

Until now parents have had to wait for up to 3 weeks for the results of this test (karyotyping), which is based on culturing the cells sampled from the amniotic fluid. Karyotyping allows examination of all the baby's chromosomes. New DNA tests have been developed that can give results in 2–3 days. These new molecular tests, using fluorescence *in situ* hybridisation (FISH) or the quantitative polymerase chain reaction (Q-PCR), search for abnormalities in specific chromosomes. Errors in chromosomes other than in those tested will not be disclosed. The abnormalities not tested for are much more rare. For example, only 4 in every 1000 babies tested will have one of these rarer abnormalities and some of these may be identified during other examinations, for example during routine ultrasound examinations.

Many parents will welcome the quicker result from a more focused test but some may be prepared to wait for the result of a test that examines all the chromosomes.

Objectives

The objectives of this study were to:

- measure the technical performance of FISH and Q-PCR tests versus karyotyping
- estimate the relative costs of molecular tests under various conditions
- establish the value to women, clinicians and others of more rapid molecular test results
- assess the cost-effectiveness of molecular tests, and consider possible changes in current testing protocols.

Design

Two-stage trial; technical performance assessed through a blinded comparison of molecular tests against the accepted gold standard (karyotyping) in a laboratory setting in the first stage; effectiveness and cost-effectiveness measured in a service setting in the second stage. Measurement of anxiety and health status of women; willingness to pay (WTP) for four stakeholder groups; and survey of UK obstetricians and midwives.

Setting

Two study sites – the catchment areas for the West Midlands Regional Genetic Laboratory and the Northern Ireland Regional Genetics Centre, Belfast.

Participants

Blinded samples: 2376 Down only molecular tests; 1576 multiplex/5-probe tests; 3952 karyotyping.

Trial: 194 women (141 intervention group; 53 control group).

WTP: 1000 general public; 141 women; 84 partners; 105 health commissioners.

Interventions (diagnostic tests)

- Molecular tests for the five most common chromosome abnormalities.
- Molecular tests for Down syndrome only.
- Karyotyping.

Main outcome measures

Technical capacity, diagnostic accuracy, diagnostic impact, patient outcome and cost-effectiveness.

Results

Technical capacity – does the test perform reliably and deliver accurate (i.e. precise) information?

FISH and Q-PCR test results are as reliable and precise as karyotyping for the five most common chromosome abnormalities.

Diagnostic accuracy – does the test contribute to an accurate diagnosis (of chromosome abnormalities)?

The ability to detect the five most common chromosome abnormalities, absolute sensitivity and specificity, are 1.00 and 1.00 for FISH and 0.9565 and 0.9997 for Q-PCR, respectively.

The ability to detect all clinically significant chromosome abnormalities, relative sensitivity and relative specificity, are 0.8605 and 0.9999 for FISH and 0.8234 and 0.9996 for Q-PCR, respectively.

Diagnostic impact – will the test replace other diagnostic tests or procedures?

Preferences of clinicians, women and other stakeholders will influence diagnostic impact.

Fifty-seven per cent of obstetricians expressed a preference for molecular tests for most patients and karyotyping for a minority; only 15% would choose both tests. The views of midwives were similar.

Most women (67%) and 54% of partners expressed a pre-test preference for molecular tests. Health commissioners were undecided. The general public expressed a preference for karyotyping (60%).

Patient outcome – does the test contribute to improved health/reduced anxiety for the patient?

Quality of life measure (EuroQol EQ-5D) demonstrated significantly increased health status linked to more rapid test results. Anxiety measure (Speilberger) exhibited similar impact.

Cost-effectiveness – does the test use improve cost-effectiveness compared to alternative interventions?

Molecular tests are less expensive than karyotyping. As a replacement within larger laboratories (> 1100 specimens per annum), Q-PCR is preferred; for smaller laboratories (< 450), FISH is preferred. Five testing regimes were assessed in terms of cost-effectiveness:

1. Molecular test and karyotyping for all women.
2. Molecular test as a replacement for karyotyping
3. Molecular test for all plus karyotyping for high-risk women.
4. Karyotyping for all plus molecular test for high-risk women.
5. Parental choice plus karyotyping for high-risk women.

Simple cost-effectiveness analysis based on the cost per case detected (all cases) demonstrates that regimes 2, 3 and 5 are more cost-effective than karyotyping and 1 and 4 are not. This pattern does not change if cost-effectiveness analysis is limited to clinically significant cases only.

Cost–utility analysis estimates a cost per quality-adjusted life-year gained of £23,542–£41,939 for regime 1; it was not possible to assess regimes 2–5 using this technique.

Regimes 2, 3 and 5 will not detect some rare chromosome abnormalities (approximately 2–4, 1–2 and 1 per 1000 women tested, respectively).

Introduction of regime 1 could increase annual UK test costs by up to £2.8 million. Regimes 2 and 3 should result in savings of up to £1.76 million per annum, and regime 5 approximately two-thirds of these savings. Regime 4 would be largely cost neutral.

Conclusions

Implications for healthcare

In the current climate, the use of prenatal testing is determined by individual clinicians, laboratories and hospitals. There is evidence of a lack of equity of provision, and of regional and local variations with regard to primary risk assessment. This may well be replicated with regard to final diagnosis if molecular tests are introduced without discussion of appropriate implementation protocols based on this report. Debate and consensus will be necessary to develop clinical protocols for introduction of molecular tests and prevent continuation of inequities and variations. Important ethical issues must not be overlooked and crucial to this debate will be the needs and wishes of parents as well as the views of other stakeholders such as scientists, obstetricians and midwives.

Recommendations for future research

It was not possible to assess the impact on quality of life and anxiety of replacing karyotyping with molecular tests for all women or selected groups of women within this study. This could be addressed ethically as tests are introduced into service and should form part of the implementation. Alternative mechanisms for delivery of test results should also be explored to optimise the advantage of faster results. There is currently little evidence of the potential impact of false-negative results on parents and on the healthcare system. If molecular tests are to replace some karyotyping tests, further research in this area is needed.

Chapter 1

Introduction and background

Introduction

In the UK, a diagnostic test for fetal chromosome abnormalities is available during pregnancy to all women who are considered to have higher than average risk factors or, less commonly, to be anxious about their pregnancy. Unlike most diagnostic tests, the purpose of this prenatal test is not to identify and treat a predefined disorder. Rather, the aim of the current diagnostic test (karyotyping) is to provide information on the presence and nature of any chromosome abnormalities in the fetus before birth, so that the parents can make informed choices about the course of the pregnancy. Termination is one option, but should the parents decide that the pregnancy will continue, birth and neonatal treatment planning can be used to optimise the outcome for the child. The information provided may also enable parents to plan for future children, whether the present pregnancy is to be continued or not.

Although karyotyping can detect a range of abnormalities, it is primarily used for detection of Down syndrome.* The test is based on an examination of fetal cells cultured either from a sample of amniotic fluid or from a sample of cells taken from the chorionic villi. Amniotic fluid is the liquid that surrounds the fetus in the uterus, and chorionic villi are tissue samples obtained from the placenta. Both amniocentesis and chorionic villus sampling (CVS) are not without some risk to the fetus. Both procedures are invasive and involve the use of a fine needle inserted through the abdominal wall under ultrasonic guidance to remove a sample of fluid or placenta cells. Amniocentesis is usually carried out in the second trimester of pregnancy (around 16 weeks); CVS is performed in the first trimester (11–14 weeks). Around 0.5–1% of amniocenteses and 1–3% of chorionic villus samples result in miscarriage.[1] CVS is the less common of the two procedures in the UK and is normally only available at centres of fetal medicine. Throughout this report, where amniocentesis samples are referred to, this term also applies to samples acquired through CVS, unless stated otherwise.

Chromosome abnormalities in the fetus

The chromosome abnormalities that karyotyping might detect in the fetus are of two types, numerical or structural. Numerical abnormalities involve the gain or loss of one or more chromosomes. This form of abnormality is termed **aneuploidy**. The presence of a single extra chromosome is referred to as **trisomy** and the presence of two or more sets of chromosomes is referred to as **polyploidy**.[2] The other main type of abnormality detected consists of structural chromosome rearrangements, and these result from chromosome breakage with subsequent reunion in a different configuration. These abnormalities include **translocations**, which involve the exchange of genetic material between chromosomes.

The most common abnormalities diagnosed prenatally are aneuploidies such as trisomy 21 (Down syndrome), trisomy 18 (Edward syndrome), trisomy 13 (Patau syndrome), and X and Y sex chromosome aneuploidies.[3] Together these account for up to 95% of live-born chromosomal abnormalities.[4] Statistics on the occurrence of the various chromosome abnormalities are complicated by the fact that certain abnormalities have such serious consequences that the chances of survival for the fetus are seriously compromised. Available data on occurrence of different abnormalities are therefore sometimes presented in relation to the number of births (with or without stillbirths), at other times in terms of pregnancies, or sometimes in relation to numbers of amniotic samples tested.

Table 1 illustrates the types of chromosome abnormalities commonly detected and their frequency in relation to birth numbers. This

* During 1999 the recognised term for the syndrome trisomy 21 became, by published convention, 'Down syndrome', replacing the earlier term 'Down's syndrome'. The newer term is used throughout this report.

TABLE 1 *Chromosome abnormalities commonly detected, their frequency and consequences*

Abnormality	Estimated incidence per 1000 births[5]	Clinical features	Proportion undergoing spontaneous pregnancy loss (%)
Unbalanced			
(i) Autosomal abnormalities			
Trisomy 13 (Patau syndrome)	0.2	High mortality in first few weeks, physical disabilities, severe mental disability	95
18 (Edward syndrome)	0.3	High mortality in first few weeks, physical disabilities, severe mental disability	95
21 (Down syndrome)	1.5	Variable (25–75 points) impairment in intelligence, short stature, cardiac anomaly 15–50% cases	80
Polyploidy	~0	Survival beyond mid-pregnancy is rare	
(ii) Sex chromosome abnormalities[a]			
Female births			
45, X (Turner's syndrome)	0.1	Intelligence normal or slightly reduced, short stature and ovarian failure/infertility	
47, XXX (triple X syndrome)	1.0	Mild reduction (10–20 points) in intellectual skills	
Male births			
47, XXY (Klinefelter syndrome)	1.0	Infertility, clumsiness and variable, but usually mild, learning difficulties in verbal skills (10–20 points) Tall stature, variable, usually mild (10–20 points) impairment in intelligence	
47, XYY (XYY syndrome)	1.0	Variable	
(iii) Others (including rearrangements)	1.0	Risk of having chromosomally unbalanced offspring	
Balanced rearrangements	3.0		

[a] *The numbers refer to the number of chromosomes present, the normal complement being 23 pairs (46 chromosomes)*

shows that the overall incidence of all chromosome abnormalities is approximately 9.1 per 1000 in newborns. Of these, the table shows that a majority of cases (6.1 per 1000 newborns) have some extra or missing chromosome material. Such chromosome abnormalities are termed 'unbalanced' and are linked to varying levels of physical and mental disability. Included in this group are trisomies 21, 18 and 13, and the numerical sex chromosome abnormalities.

The remainder of abnormalities (found in approximately 3 per 1000 newborns) are termed 'balanced' chromosome abnormalities. Individuals with these abnormalities do not usually have any symptoms. However, as **carriers** they are at increased risk of having chromosomally unbalanced offspring. In turn, these children have increased risk of miscarriage and/or the birth of children with learning difficulties and physical disabilities.

The most common chromosome abnormality is trisomy 21, commonly termed Down syndrome, where there is an extra chromosome 21. Trisomy 21 occurs in approximately 1.5 per 1000 newborns.[6] Children with this condition may suffer from learning difficulties and/or certain physical defects. For example, it is estimated that Down syndrome accounts for about 25% of all cases of moderate and severe learning difficulties in children of school age; it may also be associated with congenital heart malformations occurring in 15–50% of children with the syndrome.[5,7]

Currently in the UK, the prenatal testing process for detection of Down syndrome and other chromosome abnormalities is conducted in two steps:

- Step 1: risk assessment and case selection for step 2
- Step 2: prenatal diagnostic testing.

Effective prenatal diagnosis of chromosome abnormalities is thus dependent on both accurate

case selection and effective diagnostic testing, and steps 1 and 2 are therefore closely linked.

Prenatal risk assessment

Step 1, which consists of prenatal risk assessment, is based on a combination of epidemiological risk factors (e.g. age, family history) and test results (e.g. serum screening). Risk is normally expressed in terms of the probability of a fetal chromosome abnormality being present. Thresholds are subsequently used to select those cases that should be offered prenatal diagnostic testing. For example, if the risk is assessed to be 1 in 250 or greater, diagnostic testing may be advised.† Careful assessment is necessary because of the invasive nature of the diagnostic test procedure (i.e. amniocentesis). The likelihood of a severe fetal abnormality must be set against the possibility of loss of a 'normal' fetus.

Women with a family history of chromosome abnormalities or previous births with chromosome abnormalities are usually judged to be at risk and offered diagnostic testing. For other women, single or combined selection criteria are used to identify those who should be offered amniocentesis. These include:

- maternal age, thresholds typically set by individual clinicians in the range 35–37 years of age
- abnormal levels of serum marker(s) detected during routine screening of maternal blood
- structural fetal defects or malformations, and/or abnormal thickening of the nuchal fold detected during ultrasound scanning
- those couples where either parent is a carrier of a structural chromosome abnormality
- parental anxiety.

It is important to note that the vast majority of women undergo amniocentesis based on an assessment of risk for Down syndrome, although the diagnostic test used will in fact detect a large number of other chromosome abnormalities, as illustrated in *Table 1*.‡

Maternal age

Historically, an age of 35 years or greater was generally accepted to be an important risk factor for Down syndrome. However, if maternal age is used as the only criterion for selection for diagnostic testing, it has been found that only approximately 30% of cases of Down syndrome are identified prenatally.[7,9] This is because the majority of Down syndrome babies are born to women below the selected cut-off age. Although the absolute risk of chromosome abnormality is lower for younger women, birth rates are higher. The shortcomings of a single age-based selection criterion for diagnostic testing is now widely recognised.[10–13]

Maternal serum screening tests

A more accurate estimation of risk of Down syndrome can be achieved by combining maternal age with the result of a maternal serum screening test using various biochemical markers. Although practice varies between Trusts, routine maternal serum screening of all pregnant women (using varying combinations of biochemical markers) is now well established in the NHS.[14]

Serum screening not only enables improved prenatal diagnosis by identifying younger women who are at high risk, thereby increasing detection rates, but it may also exclude older women who are found to be at lower risk, thereby reducing the risks to the normal pregnancy associated with amniocentesis. A recent report for the Health Technology Assessment (HTA) Programme suggests that Down detection rates, using a combined maternal age and serum screening approach, and accepting a false-positive rate of 5%, range from 59% for two serum markers to 76% for four serum markers.[8] In general, selection of women for amniocentesis based on serum screening and maternal age is also reported to be more cost-effective than the use of maternal age alone.[7–9,15–18]

Nevertheless, the role of routine serum screening continues to be reassessed as the sensitivity of serum screening tests in general use is constantly re-examined.[14,19–22]

† Cut-off risk thresholds for diagnostic testing vary from hospital to hospital but are normally in the range 1:250 to 1:300. In many cases, the exact figure is decided regionally in conjunction with specialist genetics services.

‡ There is research evidence to show that the quality of information given and the understanding of patients about the nature of the results possible from karyotyping varies widely.[8] In many cases, it appears that women are being consented to a test for Down syndrome although other, albeit rarer abnormalities, will also be detected and may lead to the option for termination. Discussion of the ethical issues relating to patient consent to examination of the full karyotype is outside the scope of this report.

Ultrasound scans

Most women in the UK receive an ultrasound scan between the 18th and 22nd weeks of pregnancy (second trimester scan). This scan is used both to check the accuracy of estimates of gestational age and to examine the fetus.[19]

It has recently been argued that a mid-trimester ultrasound examination, together with maternal age, may provide a valid option for risk assessment for abnormalities such as Down syndrome, without the use of serum screening.[19] Several studies have looked specifically at the use of ultrasound scanning to identify high-risk pregnancies for Down syndrome. Earlier anomalies scanning at 11–12 weeks (first trimester) and measurement of the thickness of the translucent nuchal fold (skin at the back of the neck) may prove useful in this respect.[10] It has been demonstrated that increased nuchal translucency can be indicative of Down syndrome.[8] However, trials in a routine setting have to date given varied results, suggesting that technical issues related to the test need further work before transfer from specialist centres is feasible. Nevertheless, a diagnosis of nuchal translucency, or absolute values for the thickness of the nuchal fold, combined with age, have become an additional criterion for selection for amniocentesis in some UK centres.[19,23]

Parental anxiety

Psychological indicators are not formally part of hospital protocols for selection of women for amniocentesis. In specific instances, however, parental anxiety is known to be used as a clinical indication for amniocentesis. It appears that use of such an indicator is varied and dependent on individual obstetric consultants.[24]

Prenatal diagnostic testing

Once prenatal risk assessment has identified a pregnancy as 'at risk', the second stage in the testing process is to offer the woman a diagnostic test to identify whether a fetal chromosomal abnormality is present or not. If present, the test will identify the nature of any abnormality.

The current gold standard test used for diagnosis, **karyotyping**, is a test based on culture and examination of fetal cells obtained through amniocentesis or CVS. In the second trimester, collection of fetal cells is normally by amniocentesis and in the first trimester by CVS (and occasionally amniocentesis). Not all attempts to obtain fetal cells are successful; around 2% of all attempts to obtain amniotic fluid are unsuccessful, and between 1.3 and 5% of CVS attempts fail.[8,25,26]

Definitive diagnosis is then accomplished using conventional cytogenetic analysis of the fetal cells cultured from the amniotic fluid (or chorionic villus sample). Failure to culture the fetal cells obtained occurs in a small number of cases; the average rate reported in the UK in 1999 was 0.3% of cases.[27] In addition, there is a slight possibility of misleading results because of overgrowth of maternal cells, estimated to occur in around 0.2% of cultures.[28] Chorionic villus samples can also undergo direct analysis (without culture).

Karyotyping is a reliable and accurate test for trisomy 21 (Down syndrome). The karyotyping detection rate for Down syndrome for those samples that have been successfully cultured is 99.7%.[29] This test also allows the detection of all chromosome abnormalities.[30,31] However, the test is non-automated and involves a lengthy procedure requiring specialist scientific expertise; it is therefore slow and expensive.

From the patient's point of view, the main disadvantage of the karyotyping test is that samples must be cultured for an extended period of time. The mean time to diagnosis is currently 13.4 days (weighted mean related to number of samples processed in laboratory, range 7.3–20.2 days).[27] Large variations in the rate at which cells grow, for individual specimens, leads to a wide range of reporting times. Obstetricians usually therefore advise patients that the wait for results following amniocentesis could be between 3 and 4 weeks. This waiting for diagnostic test results can place a significant emotional burden on women and their partners.

Developments in prenatal diagnostic testing

Ideally, a prenatal diagnostic test should provide speedy as well as accurate information, and a number of methods have been proposed over the years to reduce the time required for cytogenetic diagnosis using karyotyping. These have included the harvesting of amniotic fluid cells *in situ* rather than from tissue culture flasks,[32] the introduction of specialised culture media,[33,34] and the use of lower percentages of oxygen in culture.[35] Some laboratories now record an average time to reporting of results of 7 days.[19,27] However, all these methods still require culture of cells and it has been predicted that, at best, such approaches

could only reduce analysis time to 6 days in individual cases.[36]

Three important trends predicted for future prenatal testing for Down syndrome are: (i) improved targeting of screening for identification of high-risk pregnancies; (ii) a shift from second to first trimester screening; and (iii) the introduction of molecular tests to provide more rapid results.[10]

A recent report from the HTA Programme has examined the first of these, the possibility of improving current criteria for selection for prenatal chromosome analysis.[8] The second trend is currently the subject of considerable research because moves towards earlier diagnosis of fetal abnormality (i.e. sampling of fetal cells in the first rather than the second trimester) would have obvious advantages. One approach being suggested is the wider use of CVS. However, this technique has a higher miscarriage rate than second trimester amniocentesis. Furthermore, there is a risk of inducing limb defects before 10 weeks. No formal, large-scale evaluation exists regarding the safety and efficacy of CVS relative to second trimester amniocentesis.[1,25,26] An alternative approach may be first trimester amniocentesis. This is technically possible, although the volume of amniotic fluid currently required for cell culture and karyotyping is considered to be larger than would be feasible without further increasing the risk to the fetus.[37] First trimester amniocentesis therefore remains to be fully evaluated in terms of its safety and efficacy relative to CVS and second trimester amniocentesis.

An alternative approach to earlier diagnosis of fetal abnormalities would be to reduce the time between sampling and diagnosis. Recent advances in genetic techniques mean it is now possible for laboratories to introduce new molecular methods to reduce the time between sampling and chromosome diagnosis to 24–48 hours. This would have the effect of potentially providing diagnosis up to 2 weeks earlier. These techniques also require smaller amounts of cellular material than karyotyping. Thus, the volume of amniotic fluid required for molecular testing is much smaller, and there may in future be an increased possibility of first trimester amniocentesis using molecular technology.[§]

Two main types of molecular method are currently available for prenatal diagnosis of chromosome abnormality. These are fluorescence *in situ* hybridisation (FISH) and the quantitative polymerase chain reaction (Q-PCR). These tests, which are described in greater detail in chapter 2 (see page 9), can be performed without the need to culture fetal cells.

The FISH test involves the use of chromosome-specific probes, which have fluorescent labels attached. The probes are applied to cellular preparations made without a lengthy process of fetal cell culture. Detection systems involve the use of fluorescence microscopy equipment or other similar image analysis systems. FISH is amenable to a certain amount of automation, and there are currently a few automated FISH systems on the market, but they have not yet been clinically evaluated. Some are also under development in universities.[38]

The Q-PCR test, the alternative to FISH, is based on a versatile and widely used amplification technique through which even the smallest amount of a defined DNA target can be amplified to provide quantities that are detectable and identifiable. Amplification utilises the action of an enzyme (DNA polymerase), which catalyses the rapid synthesis of new strands of DNA from an original strand using a primer. Detection requires a DNA analyser, an expensive item of equipment.

Background to the current research

Pressure from medical champions, the media, special interest groups and industry to introduce new tests before they have been fully evaluated, in terms of both their technical performance and their cost-effectiveness, has always been a challenge for those commissioning services. Currently, the NHS faces a number of difficult challenges posed by new genetic tests.[10,13,39,40] Although much of the current debate centres on the introduction of

[§] Much effort is also currently being devoted to finding new diagnostic techniques that avoid the invasive procedure of amniocentesis or CVS altogether. It is well known that fetal cells occur in maternal blood during pregnancy. The isolation of such fetal cells from maternal blood samples could provide a potential source for diagnosis of fetal chromosome abnormalities. However, the number of fetal cells in maternal circulation is very low, ranging from 1 per 10,000 to 1 per 10 million. There is no reported research yet indicating that this form of non-invasive approach can provide a reliable diagnostic test for detection of prenatal chromosome abnormalities.

entirely new diagnostic testing, made possible by genetic advances, in some instances the same technological developments can also be applied to existing diagnostic procedures. Prenatal diagnosis is one such area, with the possibility of introducing new molecular methods to provide more rapid diagnosis of fetal chromosome abnormalities.

With over 650,000 pregnancies per annum in England and Wales, piecemeal introduction of a new prenatal test technology could lead to a substantial increase in costs, especially if the new molecular tests were to be used as an 'add-on' to existing karyotyping. Furthermore, widespread roll-out of these molecular tests could also have substantial implications for the organisation of genetics services, including counselling services, and for training of laboratory staff in these new techniques. As was highlighted in a recent HTA publication, the current pattern of service in the UK already demonstrates "inequality of access to service and ... current multiple stepwise uncoordinated screening of Down's syndrome".[8]
It is important, therefore, that any widespread introduction of new molecular methods in the antenatal setting should include careful consideration of the costs and benefits of different testing strategies and their implementation.
For this to be possible, research evidence is required to support decision-making.

Aims of the research

The present project was commissioned by the NHS HTA Programme in 1997. The purpose of the study was to evaluate molecular prenatal diagnosis for Down syndrome by comparing two molecular alternatives (FISH and Q-PCR) with traditional laboratory karyotyping.

The main aim of the research project was to provide reliable evidence for the introduction of molecular methods into NHS laboratories for prenatal diagnosis of chromosome abnormalities (principally Down syndrome). Within this overall aim, the objectives of the research were:

- to establish the reliability, sensitivity/specificity and turnaround time of FISH and Q-PCR diagnosis (respectively on uncultured amniocytes and DNA extracted from amniotic fluid samples) compared with the existing culture-based test (karyotyping)
- to estimate the cost of these new techniques relative to the existing test under various conditions (e.g. different throughputs, staffing structures, diagnostic testing strategies, etc.)
- to assess the value to women and clinicians of more rapid diagnosis
- to estimate the cost-effectiveness, cost–utility and cost–benefit of these molecular tests, and to model the impact of possible changes in current testing protocols.

Table 2 provides an overview of the characteristics of the two molecular tests at the outset of the study.

Study design

The study design involved a phased introduction of the technology, with the objective of determining the extent to which routine use of molecular tests on a large scale would be reliable, effective and cost-effective compared to existing cytogenetic methods.

The design chosen was a two-stage trial with technical performance assessed in a laboratory setting in the first stage, and effectiveness and cost-effectiveness measured in a service setting in the second stage. The evaluation was conducted at two study sites (the West Midlands Regional Genetic Laboratory[¶] and Northern Ireland Regional Genetics Centre, Belfast) and the Trusts served by these sites.

The evaluation framework used in the study was based on the hierarchy formulated by Fineberg in 1977 for diagnostic technologies in general.[41] An adapted form was developed for prenatal diagnosis with the following five levels:

- **Technical capacity** – does the test perform reliably and deliver accurate (i.e. precise) information?
- **Diagnostic accuracy** – does it contribute to an accurate diagnosis of chromosome abnormalities?

[¶] Six months after commencement of the study, the West Midlands Regional Genetic Laboratory (based at the Birmingham Heartlands Hospital Trust and headed by Professor Hultén) was amalgamated and transferred to a new centralised Regional Cytogenetics Service at Birmingham Women's Hospital Trust. Stage 1 technical evaluation and Q-PCR test development laboratory research was therefore transferred to the Biological Sciences Department, University of Warwick (under Professor Hultén) during this period of disruption. Stage 2 of the study (the service evaluation of molecular tests) was conducted in the new Regional Cytogenetics Laboratory in Birmingham.

TABLE 2 *Relative advantages and disadvantages of FISH and Q-PCR at outset of study*

	FISH	Q-PCR
Anticipated advantages	Analysis completed in 24–48 hours	Analysis completed in 24–48 hours
	Able to enumerate the five chromosomes most commonly associated with numerical chromosomal abnormalities (21, 13, 18, X, Y)	Quantitative interpretation of profiles is less subjective than interpreting FISH results
	Application in cytogenetic labs well developed; all large laboratories use FISH in routine practice, although not for prenatal interphase trisomy detection	Less labour-intensive, both in terms of time and expertise, than FISH
		Q-PCR analysis already automated
Anticipated disadvantages	Interpretation of results more subjective than in quantitative Q-PCR	Application for aneuploidy diagnosis much less well developed than FISH, but DNA laboratories use Q-PCR in routine practice for other conditions
	More labour-intensive, both in terms of time and expertise, than Q-PCR	Protocols for simultaneous detection of aneuploidy for chromosomes 21, 13, 18, X, Y (multiplexing) insufficiently developed
	Consumable costs higher than Q-PCR	Capital costs considerably higher than FISH
	More difficult to automate than Q-PCR	Not able, without further development, to detect structural chromosomal abnormalities
	Not able, without further development, to detect structural chromosomal abnormalities	

- **Diagnostic impact** – does the test replace other diagnostic tests or procedures?
- **Patient outcome** – does the test contribute to improved health or reduced anxiety for the patient?
- **Cost-effectiveness** – does use of the molecular test improve the cost-effectiveness of healthcare compared to alternative interventions?

Stage 1

At the outset of the study, the developmental state of the two tests was slightly different. One test (FISH) was already developed and available as a commercial kit to test for trisomies 21, 13 and 18, and X and Y aneuploidy. The other molecular test (Q-PCR) still required further development research in order to produce a multiplex test that could detect trisomies 21, 13 and 18, and X and Y aneuploidies simultaneously. This essential development work was therefore conducted as part of stage 1 of the study.

During the first part of the study, the technical performance of the new molecular methods was assessed through an independent masked comparison of each molecular test against the accepted gold standard, karyotyping.[41–43] For Q-PCR, this included assessment of both a Down only Q-PCR test and a later multiplex Q-PCR test capable of detecting trisomies 21, 13 and 18, and X and Y sex chromosome abnormalities. For FISH, a 5-probe test, able to identify the same abnormalities as the multiplex Q-PCR test, was assessed. All amniotic samples, where sufficient residual fluid was available after allocation for karyotyping, underwent a molecular test in addition to karyotyping. For each sample, one of the molecular tests (Q-PCR or FISH) was performed because there was generally insufficient fluid for both molecular tests and karyotyping to be carried out on a single sample. The scientists performing the molecular tests were blind to the karyotyping results, and vice versa. Molecular test results were not available to clinicians during this technical evaluation stage.

A detailed costing study of the two molecular tests and of karyotyping was performed. In addition, a survey of all UK cytogenetics laboratories was carried out to identify the molecular test preferences of laboratory staff, and to gather information on training requirements, laboratory skill-mix, karyotyping test prices, access to DNA sequencers, and any planned introduction of molecular tests. In addition, a questionnaire survey of obstetricians in the two study regions was carried out to identify clinical perceptions of the benefits of faster molecular test results.

All patients with abnormal karyotyping test results, identified by both genetics services during stage 1, were followed up with a questionnaire (anonymised to the research team) to the individual's obstetrician. This questionnaire asked about the selection criteria used in each specific case, and recorded the obstetrician's retrospective test preference, supposing a molecular test had been available for this patient. A similar-sized sample of normal results was followed up in the same way.

At the end of stage 1, a preliminary evaluation of the two molecular tests was performed, taking into account costs and technical performance, in order to select one or both tests for further assessment in a routine service setting during stage 2 of the project.

Stage 2

During this second stage of the study, the benefits to women and their partners of a more rapid result were explored in a routine service setting. Although there was some experience of using FISH in a service setting in UK laboratories, including the study sites, there was no similar experience of using multiplex Q-PCR. The effective transfer of the Q-PCR test to the service setting was therefore expected to require more time and, possibly, some adaptation. Once a molecular test was established in the service setting, its technical performance could be assessed under routine service conditions during stage 2.

Although abnormal test results are important when considering the clinical benefits of a more rapid diagnostic test, these represent only a very small minority of prenatal test results (3% of cases). An important benefit to the whole group of women being tested, therefore, was expected to be associated with normal test results (97% of cases), linked to the reduced period of anxiety while waiting for test results.

During stage 2 of the study, molecular test results were made available to clinicians and patients in hospitals served by the two centres, with confirmatory karyotyping results being provided at a later date. The benefits of a molecular test result were assessed in a number of ways during this stage of the project.

The impact of molecular tests on levels of anxiety and health status of a sequential sample of women who received a molecular test was compared with a control group of women receiving only the traditional karyotyping test. Women in each group were asked to record their health status/quality of life (QoL), together with anxiety at specific points during the test wait period until 7 weeks after amniocentesis. The last measurement was therefore approximately 4 weeks after the karyotyping result. This latter measurement was made as there is evidence that anxiety may persist beyond receipt of test results.[44]

Preferences for prenatal testing were also obtained for mothers and their partners, exploring the perceived trade-off between faster results and less comprehensive information on chromosome abnormalities. The value placed on more rapid tests was also measured for a sample of women using a willingness to pay (WTP) approach. In addition, preferences and strength of preference (WTP) were measured for two further groups: (i) a representative cross-section of the general public; and (ii) a sample of health commissioners (Directors of Public Health or Consultants in Public Health Medicine). Additionally, the survey of obstetricians carried out in the West Midlands and Northern Ireland during stage 1 was extended to include all UK obstetricians.

Chapter 2

Development and technical evaluation of molecular tests

Introduction

Of the two molecular tests assessed in this project, FISH was the one with a longer history in prenatal diagnosis and at the start of the study, FISH was already playing an important role in diagnostic cytogenetics, although not for routine prenatal testing. In contrast, Q-PCR was less well developed in terms of its application for detection of fetal chromosome abnormalities, but was widely used in genetics for the diagnosis of other genetic disease. Although they differed in their history, both molecular tests represented a major technical advance, which could potentially prove to be very valuable in prenatal diagnosis.

Background to the FISH test

FISH was first introduced in clinical cytogenetics in the USA in 1988,[45] following which there was a rapid development in the technology.[46] The test was first introduced into the UK in 1991.[47] The technique uses chromosome-specific probes with fluorescent labels attached; these probes are now available in commercial kit form. Detection requires microscope systems, which may range from simple fluorescence microscopes to more expensive image analysis systems such as microscope and cooled charged-coupled device cameras. FISH is a less labour-intensive test than the current karyotyping technique. Also, although FISH is relatively expensive in terms of consumables, the capital (equipment) costs are relatively low compared with Q-PCR.

History of the FISH test in prenatal diagnosis

Early FISH studies on uncultured amniocytes made use of single chromosome-specific probes (e.g. for chromosome 21). However, these first probes (centromeric repetitive or alphoid) were found to show cross-hybridisation between certain chromosomes (for example, the two probes interacted for chromosomes 13 and 21). This discovery led to the development of different types of probe (chromosome cosmid contig probes and YACs). The use of the first of these to detect Down syndrome (trisomy 21) was successfully demonstrated in a prospective study of 500 uncultured amniotic fluid samples as early as 1994.[48] Around 2 years later, the YAC probes started to be used in the UK to provide rapid preliminary reports on samples (Lowther G, Duncan Guthrie Institute of Medical Genetics, Glasgow: personal communication, 1996). At the same time, it was demonstrated that a combination of five FISH probes could be used to examine all five chromosomes most commonly associated with chromosomal abnormalities (21, 18, 13, X, Y) in a single* multicolour FISH hybridisation experiment.[3] Subsequently, 5-probe FISH test kits were developed and marketed by commercial manufacturers.

Background to the Q-PCR test

Q-PCR, the alternative to FISH, was at the outset of the study a less well established test for prenatal diagnosis. The technique amplifies DNA by the polymerase chain reaction method using selected primers, and it is theoretically possible to amplify DNA from a single copy of the target sequence. The cost of reagents for Q-PCR is lower than for FISH, although a royalty payment has to be made for service use of Q-PCR, negotiated by the NHS with the licence holders. Quantitative interpretation of Q-PCR chromosome profiles is also less subjective than using the human eye to interpret FISH test results.[49] In addition, Q-PCR is less labour-intensive than FISH and is more easily automated; Q-PCR products can be automatically analysed using a DNA sequencer and associated software. The capital cost of this equipment is high [e.g. Applied Biosystems, approx. £93,000 (all prices in this report are based on 1998

* The hybridisation is carried out with the 13 and 21 probes applied to one area of the microscopy slide and the 18, X and Y probes applied to another area, on the same slide. Microscopy analysis is performed separately with respect to these two areas.

prices)], but many genetic services already have sequencers, or have negotiated access to sequencers run by others.

History of the Q-PCR test in prenatal diagnosis

Successful Q-PCR amplification was first described in 1991 to detect X chromosome differences,[50] at about the same time as the FISH test was being introduced into service in the NHS. The technique was then modified by Mansfield in order to be able to detect trisomies 21, 18 and the triple-X syndrome using small tandem repeat (STR) markers combined with the Q-PCR reaction.[51] A small trial (134 cases) investigating the accuracy of this technique (using the STR marker D21S11) in detecting trisomy 21 in varied samples (amniotic fluid, chorionic villus biopsy and fetal blood samples) was reported 3 years later, in 1994.[52] Subsequently, a number of small trials reported, where this technology was applied to other selected aneuploidies. Eventually, in 1995, it was shown that it is possible to use a combination of Q-PCR markers (rather than a single marker) to simultaneously diagnose trisomies 21 and 18, and X and Y chromosome abnormalities.[53,54] However, by 1997 (the start of the present study), none of these multiplex assays had been tested for prenatal diagnosis, either in large patient populations or under routine conditions in NHS laboratories.

At the outset of the study, therefore, Q-PCR had not yet reached the same level of development as FISH for prenatal testing, as illustrated in *Figure 1*.

Development of the Q-PCR test

As has been reported above, at the start of the study, although a 5-probe FISH test was available to identify trisomies 21, 13 and 18 and X, Y-related abnormalities in a single process, a similar five chromosome Q-PCR (multiplex) test was not yet available as a single test. Translational research was therefore required at the outset of this HTA project to produce a Q-PCR test that was comparable to 5-probe FISH. This research was carried out in two stages. Firstly, Q-PCR was refined and underwent scientific and preliminary technical evaluation for Down syndrome (trisomy 21) detection only, while simultaneously, a multiplex test was developed. Then, the multiplex test similarly underwent scientific and preliminary technical evaluation in the laboratory for the five most common chromosome abnormalities. By the end of this process, both FISH and Q-PCR were at a stage to undergo comprehensive technical and service evaluation (*Figure 1*).

Preliminary assessment of single Q-PCR test (for Down syndrome detection)

Before the start of the main study, Q-PCR was first assessed for the detection of trisomy 21 (Down syndrome) only at Heartlands Hospital, Birmingham (supported by the West Midlands Locally Organised Research Scheme). The material used (1430 samples) was surplus to that required for amniotic fluid cell culture. Testing of these samples using single Q-PCR demonstrated that the use of only two markers was informative

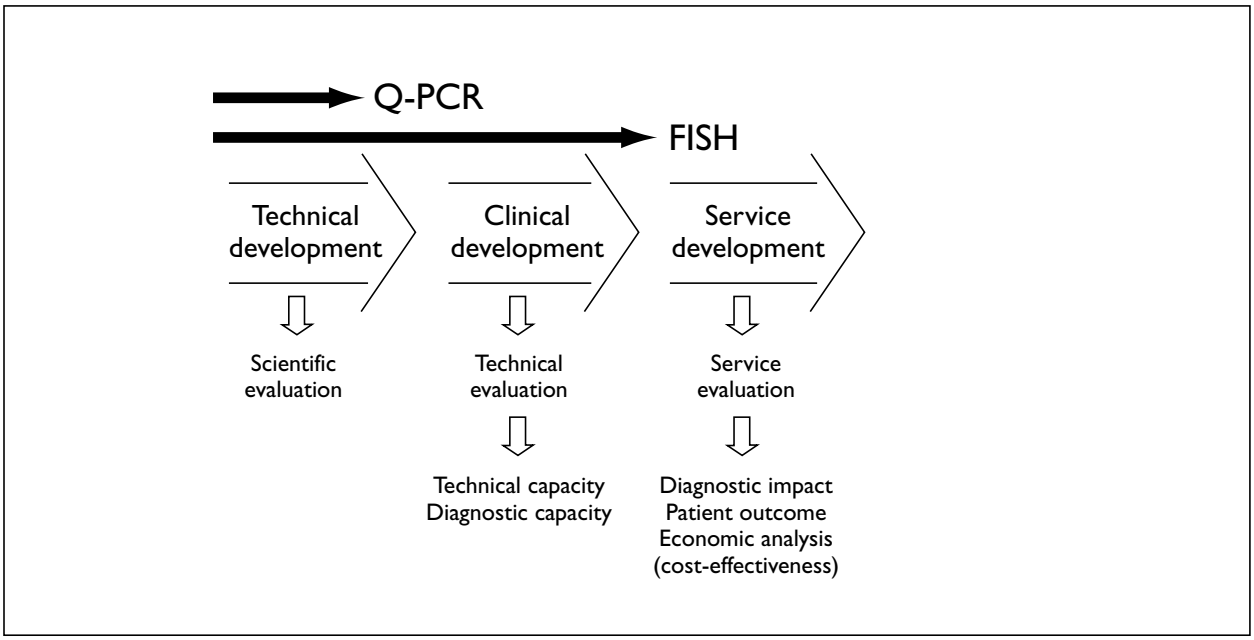


FIGURE 1 *The relative stages of development of FISH and Q-PCR at the start of the study*

in 97.6% of samples. In the remaining 2.4% of cases (34/1430), an additional marker would have been required for trisomy 21 diagnosis to be achieved. In the 1430 samples tested, only 1 cytogenetically verified trisomy 21 case fell into the group requiring an additional marker, while another 20 trisomy 21 cases were readily identified by the single Q-PCR. There were no (0/1430) false-positive or false-negative diagnostic errors per se. The only complication encountered was maternal blood contamination, which occurred in less than 1.5% of amniotic fluid samples.

At the start of this study a further 737 samples were analysed. There was no significant change in the earlier findings with these additional samples included in the analysis. Overall a total of 2139 samples were analysed, as a total 28 out of 2167 of samples (1.3%) were bloodstained and excluded. Details of the full assessment of 2167 samples, including additional samples processed as part of this study, are provided in appendix 1.

Development and preliminary assessment of the multiplex Q-PCR test

It proved more complicated than anticipated to combine markers for the detection of aneuploidies other than trisomy 21. In particular, problems were encountered in identifying which specific markers could be successfully combined and used together (multiplexed). The research team therefore had to undertake a more elaborate and staged technical development process than was initially anticipated. First, three markers were combined and used together for the detection of trisomy 21 (D21S11, IFNAR and D21S1270). Once this element of the test was optimised, other markers were added one by one. Eventually, two separate combinations of multiplexing were devised, each containing 5–6 markers. The Q-PCR test finally selected for this study used five markers for chromosome 21 (D21S11, IFNAR, D21S1270, D21S1412, D21S1411); three markers for chromosome 18 (MBP-A, MBP-B, D18S535); two markers for chromosome 13 (D13S631, D13S634); one marker (AMXY) for sexing; and one marker for the X chromosome (XHPRT).

Technical evaluation of molecular tests

Following test development and preliminary scientific evaluation, the evaluation of a diagnostic technology then focused on comprehensive technical and service evaluation of the test (*Figure 1*). In the most widely accepted hierarchy (Fineberg) for evaluating diagnostic tests,[41,55] the first two stages of such an evaluation are:

- **Technical capacity** – does the test perform reliably and deliver accurate (i.e. precise) information?
- **Diagnostic accuracy** – does the test contribute to an accurate diagnosis (of chromosome abnormalities)?

Both molecular methods (FISH and Q-PCR) were assessed in terms of these two aspects of test performance relative to karyotyping. During stage 1 of the study, the samples tested (anonymised to the research team) were drawn from amniotic fluids received at three centres: Birmingham Heartlands Hospital Trust, Birmingham Women's Hospital and the Northern Ireland Regional Genetics Centre (Belfast City Hospital Trust). The following approach was then adopted:

- a karyotyping test was carried out on each sample
- any amniotic fluid in excess of that required for cell culture and karyotyping was used to perform molecular tests (approximately 1–2 ml[†] was necessary for each test).

In the original research protocol, the plan was to perform both molecular tests, Q-PCR and FISH, in addition to karyotyping, for each sample. However, in most cases there proved to be insufficient fluid to perform both molecular tests. As a result, the research protocol was adapted to include allocation of samples to only one of the molecular tests, FISH or Q-PCR, as well as karyotyping.

Data recorded for all samples processed as part of the study included indications for referral; estimated gestational age; amount of fluid; presence of visible blood contamination; turnaround time for laboratory processing (molecular tests and karyotyping); any technical difficulties encountered in performing the test; and karyotyping and molecular test results. Molecular test results were not released to clinicians or women during the first, technical evaluation stage of the study. The only use made of molecular test results was within the laboratory, and this was left to local decision-makers. In the service laboratories, an

[†] In practice, it was found that the volume required was dependent on gestation. More fluid was needed for earlier gestation tests using FISH.

abnormal molecular test result was likely to lead to earliest possible harvesting of cell cultures and therefore a more rapid karyotyping result.

Test repeatability (the variance in a test result that occurs when the test is repeated on the same specimen) could not be assessed because insufficient amniotic fluid was available after fluid had been allocated for karyotyping and the selected molecular test.

Fineberg level 1: technical capacity of molecular tests

Technical capacity – whether a tests performs reliably and delivers accurate (i.e. precise) information – was assessed in the study laboratories for three molecular test variants: the 5-probe commercial FISH kit (AneuVysion, Vysis Ltd) and two forms of in-house Q-PCR test, a Down syndrome test and the multiplex Q-PCR test. The 5-probe FISH and multiplex Q-PCR both tested for trisomies 21, 18 and 13 and numerical sex chromosome abnormalities.

Technical difficulties were systematically recorded throughout the study. There were no reported incidences of major equipment malfunction in either stage 1 or stage 2 of the research project (described earlier, see pages 7 and 8). However, some very early failures of Q-PCR were observed in stage 2 of the study as the test entered into the routine laboratory setting, moving from the University of Warwick to the Regional Genetic Laboratory in Birmingham. These were a result of attempts to speed up reporting time by using an alternative test protocol, leaving the sample for DNA precipitation in the freezer at –70°C for 2 hours rather than at –20°C overnight. No Q-PCR amplifications were obtained for any of these samples.

The main study findings relating to the technical capacity of molecular tests are summarised in *Tables 3* and *4* for FISH and Q-PCR tests respectively. The top section of each table (labelled i) first lists the total number of samples allocated to testing using 5-probe FISH or both types of Q-PCR test (Down syndrome only and multiplex Q-PCR). Using 5-probe commercial FISH kits, a total of 859 samples were processed (400 samples at the University of Warwick and 459 in Northern Ireland). Using the Down only Q-PCR test, 2367 samples were processed (2167 samples at the University of Warwick and 200 in Northern Ireland). Following final development of the multiplex test for the other trisomies (13, 18) and numerical sex chromosome abnormalities, a further 717 samples were processed at the University of Warwick using the multiplex Q-PCR test.

When a prenatal test is carried out on a sample, there are three main types of test outcome that may be classified as a 'failure'. Firstly, the test process may itself fail. In the present instance, this might be because of a failure to culture cells from the amniocytes (in the case of karyotyping) or a failure to amplify (Q-PCR) or hybridise

TABLE 3 *Results of technical evaluation of FISH molecular tests in stage 1*

		FISH		
		Northern Ireland	**West Midlands**	**Total (%)**
i	Total samples allocated to test	459	400	859 (100)
	Karyotyping tests where no result obtained	1	1	2 (0.2)
	Molecular tests where no result obtained	54	4	58 (6.8)
ii	No. of abnormalities disclosed by karyotyping	23	15	38 (4.4)
	No. of abnormalities disclosed by molecular test	23	13	36 (4.2)
	No. of abnormalities not disclosed by molecular test	0	2[a]	2[a] (0.2)
iii	Specimens excluded from molecular tests because of visible macroscopic bloodstaining	26	Unknown[b]	Unknown[b]
	Evidence during test of maternal cell contamination in original specimen[c]	26	10	36 (4.2)

[a] *(a) 45XY, der(13:14) (p10:p10) pat and (b) 46XY del(15)(q26.1), neither expected to be detectable by FISH using the Aneuvision protocol. In both these cases clinical risk of abnormality was assessed to be high because of (i) familial history and (ii) abnormal scan result, which would imply that karyotyping would be performed*

[b] *All specimens with visible bloodstaining excluded before referral to laboratory in West Midlands*

[c] *The number of samples that showed a mixture of XX and XY cells on analysis*

TABLE 4 *Results of technical evaluation of Q-PCR molecular tests in stage 1*

		Q-PCR				
		Northern Ireland	West Midlands		Total (%)	
		Down only	Down only	Multiplex	Down only (%)	Multiplex (%)
i	Total samples allocated to test	200	2167	717	2367 (100)	717 (100)
	Karyotyping tests where no result obtained	4[a]	0	0	4[a] (0.2)	0 (0.0)
	Molecular tests where no result obtained	32	15[b]	68[c]	47 (2.0)	68[c] (9.5)
ii	No. of abnormalities disclosed by karyotyping	11	37[d]	12	48 (2.0)	12 (1.7)
	No. of abnormalities disclosed by molecular test	8	32	10	40 (1.7)	10 (1.4)
	No. of abnormalities not disclosed by molecular test	3[e]	5[e]	2[f]	8[e] (0.3)	2[f] (0.3)
iii	Estimate of total cases of maternal cell contamination	22	28	Unknown[g]	50 (2.1)	Unknown[g]

[a] *These specimens were part of the batch affected by an outbreak of bacterial contamination within the cell culture*
[b] *No more DNA left for reanalysing the cases homozygous, and therefore non-informative, for both D21S11 and IFNAR*
[c] *Includes technical problems in development pilot phase, e.g. when trying to speed up report time (see page 12) or if insufficient sample size for initial multiplexing of Q-PCR*
[d] *Q-PCR testing only for Down syndrome in this series. Five abnormalities were detected at karyotyping, none was Down syndrome*
[e] *Molecular test not capable of detecting these chromosome abnormalities, none was Down syndrome*
[f] *Both samples 47XXY (Klinefelter syndrome), the single polymorphic marker for X chromosome was uninformative*
[g] *All cases of visible bloodstaining in original amniocentesis sample excluded before forwarding to the study laboratory*

(FISH) DNA. In such a case, no test result would be obtained.

A second outcome may be that although the test process does not fail, the result obtained is inconclusive or uninformative. Such an outcome may occur in the case of Q-PCR, for example, when a second sequencer run is required to secure a reportable result but there is insufficient DNA available. In such a case also, no test result would be obtained. Finally, a test may be completed and a result produced, but there may be an incorrect result or test error.

It has long been recognised that evaluation of diagnostic tests tends to concentrate on the last of these and that the first two types of outcome (test failure and inconclusive/uninformative results) are rarely reported in the literature.[56] The present study recorded all cases where no test result was obtained, including the first two types of outcome.

The top sections of *Tables 3* and *4* (rows labelled i) indicate the number of karyotyping and molecular tests where no result was obtained. For FISH, *Table 3* shows that 58/859 molecular tests fell into this category and therefore only 801 molecular tests produced a result. For Q-PCR, *Table 4* shows that there were 47/2367 Down only tests in which no result was obtained (2320 produced a result). Similarly for multiplex Q-PCR, 68/717 tests produced no result and in the remaining 649 cases a result was obtained. Both FISH and Q-PCR therefore exhibited a relatively high incidence of test failure or inconclusive/uninformative results during the introduction period: 9.5% for multiplex Q-PCR, 6.8% for 5-probe FISH, and 2.0% for Down only Q-PCR. The comparable level for karyotyping was less than 0.2%.

Factors influencing the level of test failure or inconclusive/uninformative results varied. In the case of multiplex Q-PCR, during the development stage insufficient amniotic fluid sometimes made it impossible to repeat the test three times when the first two markers were uninformative (because the individual was not polymorphic for either of these first two markers). This accounted for many of the multiplex Q-PCR tests where no result was obtained. Another possible factor influencing Q-PCR tests, maternal cell contamination, is discussed on page 15 ('Molecular test results and maternal cell contamination'). Factors influencing the level of failed tests for FISH are discussed in more detail on page 14 ('Molecular test results and gestational age').

In the middle section of *Tables 3* and *4* (section labelled ii) information is presented, for all the

tests that were completed, on the number of specimens in which chromosome abnormalities were disclosed by the molecular test, together with the number of abnormalities disclosed in the same specimens by karyotyping. The numbers of abnormalities not disclosed by the molecular test are also listed. In all these cases, either the molecular test was not designed to disclose these abnormalities, or the test proved uninformative using a single X chromosome marker (*Table 4*).

For 5-probe FISH, *Table 3* shows that in a total of 859 samples tested, two abnormalities were not disclosed (0.2%). Both of these were undetectable chromosome abnormalities, that is abnormalities that the test was not designed to disclose. Furthermore, in both cases other factors such as familial history and abnormal ultrasound scan results would have indicated that the clinical risk of an abnormality was high before testing was initiated. In such a situation, therefore, it is likely that in the service setting a full karyotyping test would have been performed.

For the Down only Q-PCR test, *Table 4* shows that in a total of 2367 samples tested there were eight chromosome abnormalities that were not disclosed by the molecular test. However, once again these abnormalities were all undetectable – abnormalities that could not be disclosed by the Down only form of the Q-PCR test. Therefore, with respect to diagnosis of Down syndrome, the first stage study results demonstrated that both FISH and Q-PCR are able to provide identical results to those obtained through the slower process of karyotyping.

For the 717 multiplex Q-PCR tests performed in the first stage of the study, *Table 4* shows there were two abnormalities that were not disclosed (1.4%). Both of these were 47XXY (Klinefelter syndrome) and cases where the single polymorphic marker for the X chromosome was uninformative. The issue of levels of undetectable abnormalities is discussed further on page 16 ('Fineberg level 2: molecular test diagnostic accuracy') and page 21 ('Relative sensitivity and specificity of molecular tests').

Finally, in the bottom section of *Tables 3* and *4* (section labelled iii) the numbers of amniocentesis samples excluded from molecular testing because of visible (macroscopic) bloodstaining are presented, where recorded. These include samples allocated to FISH and those allocated to Q-PCR. In the West Midlands, all visibly bloodstained samples were excluded before they were sent to the laboratory for processing. Overall, for the Q-PCR Down only test the estimated level of samples with maternal cell contamination in both laboratories was 2.1%. In the West Midlands, visible maternal blood contamination was observed in 1.3% of amniotic fluid samples (see appendix 1). For FISH, the number of samples (4.2%) that subsequently demonstrated evidence of maternal cell contamination during molecular testing in both centres is also listed in *Table 3*, section iii. FISH samples in Northern Ireland recorded a higher pre-test level of 5.7% (26/459). The issue of maternal cell contamination and its impact on test results is considered further on page 15 ('Molecular test results and maternal cell contamination').

Molecular test results and gestational age

The main factor affecting the reported high level of inconclusive results for multiplex Q-PCR was the lack of sufficient amniotic fluid to repeat the test when initial markers were uninformative, as was explained on page 12 ('Fineberg level 1: molecular test technical capacity'). However, for Down only Q-PCR and FISH tests the likelihood that no result would be obtained, as reported in *Tables 3* and *4*, also appeared to be linked to geographical location. Such results seemed to be mainly concentrated in Northern Ireland, where 54/58 of the failed FISH tests and 32/47 of the failed Down only Q-PCR tests were observed.

Table 5 presents a breakdown of failed tests by gestational age for FISH tests carried out in both centres during the first stage of the study. This table illustrates that failed tests across the two sites were disproportionately concentrated in the early gestational samples (11–13 weeks) and also the very late samples (23–36 weeks) encountered in Northern Ireland.

It appears therefore that the differences observed in the level of failed tests in the two sites may be in part attributable to differences in the pattern of testing in the two geographical areas. The Regional Genetics Centre has a much higher proportion of early amniocentesis samples; in fact, the laboratory probably has more experience of early testing than most other UK laboratories. The laboratory also deals with a higher proportion of late amniocentesis samples than is the case in the West Midlands.

In early amniocentesis, samples may contain a lower proportion of fetal cells, with fewer cells available for analysis; and in late amniocentesis,

TABLE 5 *FISH failed tests versus gestational age compared with karyotyping during stage 1 of the study*

	Failed tests, West Midlands			Failed tests, Northern Ireland		
Gestation (weeks)	**Number of samples (%)**	**FISH**	**Karyotyping**	**Number of samples (%)**	**FISH**	**Karyotyping**
11–13	4 (1)	0	0	116 (25)	42	1
14–16	165 (41)	0	0	208 (45)	4	0
17–19	216 (54)	0	0	97 (21)	0	0
20–22	15 (4)	4	1[a]	24 (5)	1	0
23–25	0	0	0	6 (1.5)	2	0
26–28	0	0	0	1 (0.5)	0	0
29–36	0	0	0	6 (1.5)	5	0
Total	400 (100)	4	1	458[b] (100)	54	1

[a] *One sample failed FISH and karyotyping*
[b] *Gestational age data missing in one case*

samples may contain a larger proportion of cells that are unusable because they are no longer alive. Both types of sample may therefore be more liable to test failure with a molecular test. High levels of inconclusive results for late gestation samples have also been noted in other reports.[57]

Although the results for failed tests in relation to gestational age are illustrated here for FISH, the samples processed using Q-PCR in Northern Ireland were drawn from the same patient population and therefore would be expected to have a similar profile.

Molecular test results and maternal cell contamination

Maternal cell contamination may also be a factor in test failure or occurrence of inconclusive results for molecular tests, particularly for Q-PCR. During amniocentesis it is possible for the amniotic fluid sample (or chorionic villus sample) to become contaminated with maternal cells (usually maternal blood). Maternal contamination may in some cases result in visible (macroscopic) bloodstaining; in other cases it may not be visible to the naked eye. Contamination is more likely in first trimester testing when CVS is also more common. The degree of contamination is also dependent on the skill of the clinician performing the procedure.

For the karyotyping test, maternal cell contamination does not appear to have a significant effect; it is reported to produce a misleading result in around 1 per 500 cases.[58] However, molecular tests, particularly Q-PCR, appear to be more sensitive to maternal contamination.

In the West Midlands, during the multiplex Q-PCR development phase, all samples were examined by eye before processing, and any sample with visible bloodstaining was excluded from testing. In Northern Ireland all samples were processed, regardless of visible bloodstaining. In both centres all samples were centrifuged at the start of processing.

A detailed analysis was carried out of the 200 samples in Northern Ireland processed using Q-PCR. Samples were divided in terms of the level of bloodstaining into 'heavy' (i.e. high level of macroscopic bloodstaining visible to naked eye), 'moderate' (i.e. some macroscopic bloodstaining visible in sample), and 'slight' (i.e. following centrifuging of the sample at 13,000 rpm for 5 minutes, some visually perceptible bloodstaining of the cell pellet). *Table 6* illustrates the levels of bloodstaining recorded for samples of different gestational age. Overall, 57 (28.5%) samples were identified to have some level of bloodstaining. The majority of these were first trimester amniocentesis samples (33/57); 41.7% of first trimester samples exhibited some bloodstaining compared to only 19.8% of second trimester samples.

Analysis of molecular test results for these specimens from Northern Ireland demonstrated that the presence of bloodstaining did not necessarily result in a failed test or an inconclusive result. The pattern of failed or inconclusive tests was examined in relation to levels of bloodstaining and gestational age of samples. It was found that the total number of failed or inconclusive results (21 or 10.5% of samples) was lower than the number of bloodstained specimens (57 or 28.5% of samples).

TABLE 6 *Visible bloodstaining for sample of 200 specimens tested by Q-PCR molecular method versus gestational age*

Gestation	Total number (%)	Level of bloodstaining				Total no. of bloodstained (%)
		None (%)	Slight (%)	Moderate (%)	Heavy (%)	
First trimester	79 (100)	46 (58.2)	23 (29.1)	8 (10.1)	2 (2.5)	33 (41.8)
Second trimester	121 (100)	97 (80.2)	12 (9.9)	10 (8.3)	2 (1.6)	24 (19.8)
Total samples	200 (100)	143 (71.5)	35 (17.5)	18 (9.0)	4 (2.0)	57 (28.5)

A similar pattern was evident in both gestational groups; 20.2% with a failed test or an inconclusive result versus 41.7% bloodstained for first trimester samples and 4.1% versus 19.8% for second trimester samples. Furthermore, the level of bloodstaining (i.e. heavy, moderate, slight) did not appear to be directly related to the likelihood of test failure or an inconclusive result. For samples with no visible bloodstaining, 10/143 (7%) produced inconclusive test results; 7/35 (20%) of slightly stained samples produced a failed test or inconclusive results; 4/18 (22%) of moderate stained samples; and 0/4 (0%) of heavily stained samples.

Taken together, these results suggest that Q-PCR appears to be especially prone to test failure or inconclusive results for first trimester amniocentesis samples, regardless of the presence or absence of visible bloodstaining. Thus, it would appear that laboratory protocols for molecular tests on samples with visible bloodstaining, including staining in the pellet, may need further refinement.

Fineberg level 2: diagnostic accuracy of molecular tests

The previous section has discussed study findings relevant to the 'technical capacity' of both molecular tests (i.e. does the test perform reliably and deliver accurate information?). The hierarchy used for evaluation of diagnostic tests distinguishes this from the question of 'diagnostic accuracy' (i.e. does the test contribute to an accurate diagnosis of chromosome abnormalities?).

When assessing 'diagnostic accuracy', in most instances the test being assessed is compared with another established (gold standard) test that demonstrates 'information equivalence'. That is, there is an expectation that both tests could in principle provide the same information. Diagnostic accuracy for such tests can then be measured in terms of simple sensitivity and specificity and/or receiver operating characteristic (ROC) curves.[56] However, in the present case the situation is fundamentally different. The tests being compared (karyotyping and molecular FISH/Q-PCR tests) are not inherently equivalent in terms of the diagnostic information they are structured to provide. Thus, it is necessary to first consider the degree and type of any such lack of 'information equivalence', before examining the diagnostic accuracy of these tests.

The molecular tests evaluated in this study are not directly comparable with karyotyping in that they are not able to detect all the chromosome abnormalities that karyotyping can identify. In a service setting, were all women to be offered a molecular test as the only diagnostic test conducted there would therefore inevitably be some abnormalities that would remain undetected until birth or later. Conversely, the traditional karyotyping test currently detects some abnormalities that have little clinical significance, either because the prognosis is unpredictable or because the abnormality will have little significance in terms of long-term outcomes for the fetus or child. In addition, it is important to note that, using current selection criteria, most women are offered amniocentesis on the basis of risk factors associated only with Down syndrome. In relatively few instances, therefore, will women have been identified as being at risk of other chromosome abnormalities, and many women are therefore under the impression that they are consenting to a test for Down syndrome.[59,60]

This lack of information equivalence between a molecular test[‡] and karyotyping will inevitably impact on any statement of diagnostic accuracy

[‡] Two forms of molecular test have been discussed – one for Down syndrome only and one structural to detect trisomies 21, 18 and 13 and numerical sex chromosome abnormalities.

because one crucial element of diagnostic accuracy is the number of clinically significant cases missed. It is important to assess the likely size of any such effect were molecular tests to be introduced into routine use, taking into account two factors: certain women should always be offered karyotyping because of other risk factors, and consideration of the clinical significance of any cases missed.

Molecular tests as a replacement for karyotyping

In order to assess the likely incidence of undetectable abnormalities were molecular tests to completely replace karyotyping, all karyotyping test results for 1997 and 1998 in the two participating laboratories were extracted (7451 from the West Midlands and 972 from Northern Ireland) and a detailed assessment of the chromosome abnormalities detected by karyotyping carried out. This analysis first identified the number and types of abnormalities that would have remained undetected had a 5-probe FISH or multiplex Q-PCR molecular test been the only diagnostic test conducted (appendix 2). These undetected abnormalities were then further assessed to identify those that would be considered clinically significant.

Table 7 summarises the 34 clinically significant abnormalities (0.4% of the total 8423 samples tested) that would have been missed if a 5-probe FISH or multiplex Q-PCR molecular test had been used rather than karyotyping. The cases included in the 'clinically significant' category were those where the laboratories in the study judged that, based on the karyotyping results, parents would be advised that there was a probability of fetal disability for the chromosome abnormality (risk figures varying with the specific chromosome rearrangement). Cases judged to be non-clinically significant were those where abnormalities were inherited or balanced rearrangements and therefore very unlikely (although no absolute certainty in these cases) to have an associated risk (appendix 2).

Clinical significance

It is important to note that the distinction between clinically significant and non-significant cases is not hard and fast. Some scientists would include fewer cases in the clinically significant category, some might include more.

There would be no uncertainty about the non-mosaic[§] abnormalities in *Table 7* (representing 17 out of the abnormalities listed). In these cases scientists are highly likely to agree that the chromosome abnormality will cause severe fetal abnormalities and, for the non-mosaic abnormalities listed in *Table 7,* there will have been discussion of termination of these pregnancies by the obstetricians, parents and clinical geneticists. Furthermore, it should be remembered that in some of these potentially missed non-mosaic chromosome abnormalities the fetus would probably not survive to birth.

In contrast, there is less certainty about mosaicisms (marked with an asterisk in *Table 7*). Where mosaicism is detected prenatally it is not possible to accurately predict the likely degree of effect on the child for two reasons: (i) the degree of mosaicism detected *in vitro* may not accurately predict the levels *in vivo*; and (ii) an individual who has mosaicism can vary from normal or near normal to having full-blown features of, for example, trisomy (multiple anomalies and severe mental disability). The laboratories concerned indicated that they would have reported these cases to parents and obstetricians as ones with a degree of uncertainty. This degree of uncertainty with regard to the possible phenotypic outcomes in mosaicism is sufficient for parents and their obstetricians to consider termination. Mosaicisms represented 50% (17/34) of abnormalities listed in *Table 7.*

It is not possible to report the number of cases in *Table 7* in which a decision was made to terminate the pregnancy. The West Midlands laboratory audit system is only 80–90% complete in respect of pregnancy outcome, and therefore the laboratory did not consider it could release this partial data.

Targeted case selection

Some of the women included in *Table 7* will have been identified, before amniocentesis, as being at risk of a chromosome abnormality other than Down syndrome, perhaps because of previous pregnancy, abnormal ultrasound scan or family history (including known parental carriers of a structural chromosome abnormality). It is highly likely that for molecular tests in use in a routine service setting these individuals would be offered

[§] Mosby's Dictionary defines mosaicism as "a condition in which an individual or an organism that develops from a single zygote has two or more cell populations that differ in genetic constitution".[2]

TABLE 7 *Clinically significant chromosome abnormalities that would have been undetected during 1997 and 1998 if molecular techniques only had been used*

West Midlands	Northern Ireland
46,XX,der(5)t(5;8)(p13.3;q24.1)mat	46,XY,trp(9) (pter->p13::p13->p24::p24->qter)
94,XXXX,+22,+22	mos 47,XX,+22[42]/46,XX[8]*
46,XX,invdup(8)(pter;;p11p23;;qter)de novo	46,XX,del(13)(pter->q31:)
46,XX,del(5)(q?33q?34)de novo	46,XY,dup(4)(pter->q33::q32->qter)
47,XY,+mar de novo.ish der(19)(wcp19+,D1/5/19Z1+)/46,XY	
47,XY,+mar de novo.ish der(22)(D14/22Z1+)/46,XY	
47,XX,+dic(15)(q11.2)de novo	
45,X,t(1;2)(q12;p25)de novo	
46,XY,del(2)(q12.2q13)de novo	
46,XY,del(15)(q26.1)de novo	
46,XY,ins(11;21)(q23;q21q22.3)t(11;21)(q22.3;q25)de novo	
47,XX,+mar.ish der(22)(D22Z1+,wcp22+)	
46,XY,add(4q),t(8;17)(p21;q21)de novo	
47,XX,+20/46,XX*	
47,XY,+8/46,XY*	
47,XX,+20/46,XX*	
47,XX,+8/46,XX*	
46,XY,add(10q) de novo	
47,XY,+der(8) (qter->p21:) de novo/46,XY*	
47,XX,+dic(15) q11.2)de novo /46,XX*	
47,XY,inv(3)(p11.2q23)pat,+8/46,XY,inv(3)(p11.2q23)pat*	
47,XY,+16/46,XY*	
47,XY,+mar/46,XY*	
47,XX,+8/46,XX*	
45,X/46,XX*	
45,X/46,XX*	
45,X[2]/46,XX[36]*	
45,X[13]/46,XY[12]*	
Mos 47,XXY/46,XY*	
45,X/46,XX*	

* *Autosomal or sex chromosome mosaics*

karyotyping, either as a replacement or as an adjunct to the molecular test. Case selection based on individual risk factors will reduce the numbers of abnormalities not detected. This is examined in greater detail in the next section.

Finally, the number and variety of 'missed' chromosome abnormalities illustrated in *Table 7* are for the West Midlands and Northern Ireland and will vary in other locations according to selection of cases for amniocentesis.

Sensitivity and specificity of molecular tests

Test diagnostic accuracy is usually presented in terms of sensitivity and specificity, or the ability of different tests to detect positive (abnormal) and negative (normal) results correctly. Sensitivity is defined as the ability of a test to detect disease when disease is present (see next page), thus it represents the probability of a true-positive result set against the probability of the disease being present.

Specificity (see below) is the ability of the test to discriminate a true-negative result in the absence of disease. To evaluate a new test, the results produced with this test are compared with the best test currently available to detect the disease or condition, the gold standard. The gold standard is thus taken as the absolute measure of the presence or absence of the disease or condition. Karyotyping is commonly held to be the gold standard test for the detection of prenatal chromosome abnormalities.

Sensitivity = number of true-positives divided by number with a chromosome abnormality (i.e. true-positives plus false-negatives).

Specificity = number of true-negatives divided by number who do not have a chromosome abnormality (i.e. true-negatives plus false-positives).

As discussed in the previous section, one of the key issues for the assessment of molecular prenatal tests is that like is not being compared with like. The molecular tests evaluated cannot detect the rarer abnormalities that karyotyping can detect. Thus, the performance of molecular tests can be viewed in two ways, expressed and calculated in terms of two different types of sensitivity/specificity measure:

- **Absolute performance** – the capability to detect all the abnormalities that the new test is capable of detecting compared to the gold standard for these same abnormalities.
- **Relative performance** – the ability to detect any abnormality within the range of all possible abnormalities that can be detected by the gold standard.

The gold standard, karyotyping

Detection error rates for the gold standard test, karyotyping, are typically quoted as being in the range 0.1–0.6%,[58] with the vast majority of errors reported to be incorrect sex assignment, as a result of either maternal cell contamination or laboratory error, but occasional missed trisomies are also noted.[¶] In addition, test failure (as opposed to error), arising from a failure to culture cells from the amniocytes, is relatively rare. For example, in Northern Ireland over a period of 10 years, 5429 amniotic fluid samples were received by the Regional Genetics Centre, and of these 62 (1.14%) failed to culture (Trainor B, Northern Ireland Regional Genetics Centre, Belfast: personal communication, 2000). Thus for karyotyping, the overall error and failure rate is low, with approximately 4–14 cases per 1000 (0.4–1.4%) expected to fail and up to a further 6 cases per 1000 (0.6%) to misdiagnose.

Molecular test performance in both stages of study

During the second stage of the study, samples where no result was obtained, either through failure of the test process or because results were inconclusive or uninformative or because of test error, continued to be recorded as for stage 1, for both FISH and Q-PCR (see page 12, 'Fineberg level 1: molecular test technical capacity). During this part of the study, the 5-probe FISH test was performed under routine service conditions in the Northern Ireland Regional Genetics Centre, and the multiplex Q-PCR test under routine conditions in the West Midlands Regional Genetic Laboratory and Consultancy Services.

In Northern Ireland, the same test protocol was used for FISH as had been used in the first stage of the study. From the 67 samples processed by the laboratory during stage 2, two (3%) proved to have insufficient fluid (< 1 ml) to attempt FISH and two (3%) failed to hybridise. Results were therefore produced for a further 63 samples. As in the first stage of the study, all FISH results were concordant with karyotyping test results. In total, 902 FISH test results were compared with karyotyping results during stages 1 and 2 of the study, and there were no false-positives or false-negatives.

In the West Midlands Regional Genetic Laboratory, multiplex Q-PCR was introduced into the service setting during the second stage of the study and a series of Q-PCR tests were performed under routine service conditions. Of the 78 Q-PCR samples processed, two samples (2.6%) did not have enough residual DNA to attempt a multiplex Q-PCR molecular test in addition to karyotyping, two samples (2.6%) failed to amplify DNA. Results were therefore produced for 74 samples.

The multiplex Q-PCR results for 73 of these samples were concordant with karyotyping. In the remaining case, however, the molecular test gave a positive result (Klinefelter/47XXY) that was not confirmed at karyotyping. There was not enough DNA remaining to re-process this sample

[¶] Subtle structural aberrations, low-grade chromosome mosaicism and fragile X were excluded by those reporting, as karyotyping is not an effective test for these and similar abnormalities.

through the sequencer, as would be normal practice if molecular tests were to be the only test conducted.** Subsequent audit of the laboratory procedure for this case identified the possible cause as contamination. The lanes on either side of the sample producing this result contained samples from male fetuses, suggesting one possible source of contamination.

In the West Midlands samples, a total of 3318 Q-PCR test results (Down only and multiplex) were compared with karyotyping during stages 1 and 2. All but one of these results (Klinefelter/47XXY, as discussed in the previous paragraph) were concordant with karyotyping. In addition, during stage 1 there were two cases, both Klinefelter/47XXY, where the single polymorphic marker for the X chromosome was uninformative.

Absolute sensitivity and specificity of molecular tests

The data collected on the technical performance of FISH and Q-PCR in both stages of the study were first used to calculate absolute sensitivity and specificity values (i.e. values based solely on those chromosome abnormalities that the molecular tests are capable of detecting). *Table 8* shows overall findings based on the 902 FISH test results and 3318 Q-PCR test results obtained in the first and second stages of the study. Absolute sensitivity and specificity figures are presented for FISH and Q-PCR relative to karyotyping for all the tests completed.†† The absolute sensitivity of FISH relative to karyotyping is 1.00 (for the conditions the test is designed to detect), whereas the absolute sensitivity of Q-PCR is calculated to be slightly lower, at 0.9565, indicating the occurrence of one false-positive result. Similarly, the specificity of Q-PCR is very slightly lower than that of FISH because of the two uninformative results.

The calculation of sensitivity and specificity in *Table 8* excludes cases where DNA amplification or hybridisation failed (molecular tests) or cell culture failed (karyotyping), and the level of these is listed in the first column. The failure rates presented exclude all samples with visible blood contamination; these were excluded from processing for Q-PCR and excluded before calculation for FISH (to ensure parity). Broadly similar figures are observed for Q-PCR and FISH, with failure levels of 3.13% and 3.66%, respectively. The reported failure rate for karyotyping (mean 1.09%) was lower than the molecular test failure rate. These results concur with the results from a large series of 5348 pregnancies where sensitivity was found to be 99.6%, specificity 99.98% and inconclusive results occurred in 2.8% of these cases.[61]

It should be borne in mind that in some instances a molecular test failure may be regarded as an effect of the study protocol. In a trial setting, where it was vital to also provide a routine diagnostic test result for the patient, sufficient amniotic fluid had first to be allocated for karyotyping before the molecular test could be attempted. In cases where only 1 or 2 ml of fluid remained, DNA occasionally failed to amplify (Q-PCR) or hybridise (FISH). This type of failure would remain a possibility if a combined testing strategy (i.e. molecular tests and karyotyping)

TABLE 8 *Absolute sensitivity and specificity of tests for chromosome abnormalities*

		Completed tests	
Test	**Failed to complete test[a] (%)**	**'Absolute' sensitivity**	**'Absolute' specificity**
FISH	3.13[b]	1.00	1.00
Q-PCR	3.66[b]	0.9565	0.9997
Karyotyping	1.09[c]		

[a] *Including all tests conducted during stages 1 and 2 where no result obtained*
[b] *Samples with visible maternal cell contamination excluded*
[c] *Based on West Midlands and Northern Ireland data*

** It is assumed that there would be sufficient amniotic fluid to repeat a molecular test if karyotyping were not used. If both molecular test and karyotyping were to be conducted, a repeat would not be necessary.

†† For the purposes of this calculation the two uninformative results were treated as false-negatives because a positive result was not produced for an abnormality that was detectable by the test. In practice, were sufficient amniotic fluid left after the first analysis the test would have been repeated with additional markers.

were to be adopted. It would, however, be removed if molecular tests were to replace karyotyping, in which case all the amniotic fluid would be available for the molecular test.

Relative sensitivity and specificity of molecular tests

In order to estimate the relative sensitivity and specificity of FISH and Q-PCR tests (i.e. values based on all chromosome abnormalities detectable by the gold standard, not just those detectable by the new test), it was necessary to first establish the likely incidence of abnormalities that the molecular tests would not detect but which would be detected by karyotyping.

It is important to note in this respect that currently a significant number of fetal chromosome abnormalities are not detected before delivery or miscarriage because positive cases are not necessarily selected for amniocentesis and karyotyping based on the primary risk assessment process. Selection criteria (as discussed in chapter 1) include serum triple test result, maternal age, increased nuchal translucency, fetal abnormality detected on ultrasound scan or a combination of these. It is also important to note that most women are offered amniocentesis on the basis of risk factors associated with Down syndrome alone.

During stage 1 of the study, a preliminary analysis was undertaken to estimate the number of clinically significant abnormalities that might be undetectable if molecular tests were to replace karyotyping (*Table 7*). This analysis was based on all tests carried out by the two study laboratories in 1997–98. In the second stage of the project, this analysis was extended further, to include all tests carried out over a substantially longer period; data for 11 years (1988–99) from the West Midlands and for 10 years (1989–99) for Northern Ireland. *Table 9* presents data on the number of different types of chromosome abnormalities detected by the two study laboratories during these periods.[29,62] In this table, the category 'other abnormalities' represents the total number of all chromosome abnormalities (clinically significant and non-significant) that molecular tests are currently not designed to detect. The total prevalence figure for these abnormalities represents 1.1% of all test results in the West Midlands and 1.7% in Northern Ireland.

For the larger laboratory (West Midlands), additional information was available that enabled the cases shown in *Table 9* to be further divided into the following two subgroups:

- **Low risk** – those cases selected for amniocentesis based solely on assessment of risk for Down syndrome during antenatal screening,‡‡ risk factors used may include maternal age and positive serum test. In rare cases amniocentesis may be requested because of maternal anxiety.

TABLE 9 *Abnormalities detected by karyotyping over an extended year period in the two health regions participating in the study*

Period of audit	Prevalence[a] (West Midlands) 1988 to 1999			Prevalence[a] (Northern Ireland) 1989 to 1999	Total
	Low risk[b]	High risk	Total		
Total cases	30,861	2,193	33,054	5,429	38,483
Trisomy 21[c]	390 (1.3%)	74 (3.4%)	464 (1.4%)	96 (1.8%)	560
Trisomies 13 and 18	63 (0.2%)	71 (3.2%)	134 (0.4%)	51 (0.9%)	185
Sex chromosome abnormalities	65 (0.2%)	22 (1.0%)	87 (0.3%)	35 (0.6%)	122
Other abnormalities	245 (0.8%)	133 (6.1%)	378 (1.1%)	93 (1.7%)	471
Total abnormalities	763 (2.5%)	300 (13.7%)	1,063 (3.2%)	275 (5.1%)	1,338

[a] *Based on those tested, rates are therefore dependent on selection criteria for amniocentesis*
[b] *Also included in this category, as were* de novo *abnormalities*
[c] *If the molecular test for Down syndrome only is used, triploidy would appear to be Down syndrome. If a test for trisomies 18, 13 and 21 was used, triploidy would be identified. Figures for numbers of triploidy are included as appropriate*

‡‡ Abnormalities detected by ultrasound would be assigned to the high-risk group.

- **High risk** – those cases where pregnancies were identified as having an additional risk, arising from past reproductive history, in particular family history indicative of chromosome abnormalities (including either parent being a carrier of a structural chromosome abnormality) or abnormal ultrasound scan. In these instances an expert judgement was made that these women would inevitably be offered karyotyping.

If it is assumed that samples from all high-risk pregnancies will have karyotyping requested, then *Table 9* demonstrates that the West Midlands prevalence of 'other abnormalities' that would not be detected by molecular tests will be reduced from 1.1% to 0.8%.

The West Midlands data further divided results into those where the abnormality was judged to be clinically significant and those where the result was of little clinical significance (or the clinical significance is unknown). Of the 245 pregnancies in the low-risk group identified to have 'other abnormalities', only 84 were judged to be clinically significant. For the West Midlands, this produces a final likelihood figure for significant chromosome abnormalities that molecular tests, once introduced into a service setting, would not be able to detect of 2.7 per 1000 (0.27% or 84/30,861). It should also be borne in mind that data presented in *Table 7* indicated that approximately 50% of these cases may be mosaicisms. In these pregnancies it is more difficult to accurately predict the likely impact of the chromosome abnormality (see page 17, 'Clinical significance').

The estimated 2.7 clinically significant cases per 1000 for low-risk women contrasts with the much higher figure of 26 cases per 1000 (58/2193) for the high-risk group, which would be undetectable were molecular tests used. For both high- and low-risk groups, however, the majority of 'other' abnormalities (236/378) are judged to be non-significant. In *Table 10*, figures presented in *Table 9* are converted to an expected number of abnormalities occurring per year for each of the regions in the study.

Using the values in *Tables 9* and *10*, the likely annual incidence of abnormalities that would remain undetected for the current case selection can be calculated for two potential molecular testing regimes. In both regimes it is assumed that all high-risk women would be offered karyotyping but the remainder of women, those categorised as low risk, would be offered either (i) a Down only molecular test, or (ii) a multiplex Q-PCR or 5-probe FISH test as a replacement for karyotyping.

Table 11 presents, for the West Midlands,[§§] the number of cases of clinically significant chromosome abnormality that would have remained undetected annually (until the end of the

TABLE 10 *Estimated annual rates of abnormalities detected per centre for all those selected for amniocentesis*

	West Midlands	Northern Ireland	Total[a] (%)
Average throughput per annum	3005	543	3548
Average total abnormalities detected	97 (100%)	28 (100%)	125 (100%)
Trisomy 21	42 (43.3%)	10 (35.7%)	52 (41.6%)
Trisomy 18	9 (9.3%)	4 (14.3%)	13 (10.4%)
Trisomy 13	4 (4.1%)	1 (3.6%)	5 (4.0%)
Numerical sex chromosome abnormalities	8 (8.2%)	4 (14.3%)	12 (9.6%)
Total 'other' abnormalities	34 (35.1%)	9 (32.1%)	42 (33.6%)
Clinically significant 'other' abnormalities	13 (13.4%)	2 (7.1%)	15 (12.0%)
Average total clinically significant abnormalities	75 (77.3%)	21 (75%)	96 (76.8%)

[a] *Columns for West Midlands and Northern Ireland are rounded to the nearest whole number; values in the totals column may reflect these rounding errors*

[§§] These figures are not calculated for Northern Ireland as data were not available on the relative proportion of high-risk cases likely to be selected for karyotyping.

TABLE 11 *Significant clinical defects that would not be detected using molecular tests per annum in West Midlands if karyotyping performed for all high-risk cases*

Testing regime	Clinically significant defects not detected per annum in West Midlands (%[a])
(i) Molecular test for Down only (trisomy 21)	21 (30.7)
(ii) Molecular test for trisomies 21, 18, 13 and X, Y abnormalities	6 (9.2)

[a] *As percentage of total clinically significant cases detected*

TABLE 12 *Relative sensitivity and specificity of molecular tests compared to gold standard karyotyping for detection of all chromosome abnormalities*

Molecular test regime compared to karyotyping	'Relative' sensitivity	'Relative' specificity
(i) FISH test for Down only (trisomy 21)	0.6478	0.9999
(i) Q-PCR test for Down only (trisomy 21)	0.6196	0.9996
(ii) FISH test for trisomies 21, 18, 13 and X, Y abnormalities	0.8605	0.9999
(ii) Q-PCR test for trisomies 21, 18, 13 and X, Y abnormalities	0.8234	0.9996

pregnancy or beyond) if all low-risk cases had either (i) a Down only molecular test, or (ii) a 5-probe/multiplex molecular test and all high-risk cases had karyotyping.

Using the data in *Tables 11* and *8* it is possible to calculate the relative sensitivity and specificity of the FISH and Q-PCR tests for low-risk cases under both testing regimes. *Table 12* presents these values for testing regimes (i) and (ii) compared to the current gold standard (karyotyping).

Table 12 demonstrates that the relative specificity for all four tests is very high, comparable to absolute specificity figures in *Table 8*. Relative sensitivity figures are, however, significantly lower than absolute sensitivity for the Down only test (0.62–0.65), as would be expected because other abnormalities will not be detected. For the 5-probe/multiplex tests, the relative sensitivity values are higher, at 0.86 and 0.82 respectively. All these values are based on the likely incidence of clinically significant abnormalities that molecular tests would not detect were all high-risk cases to be offered karyotyping and were case selection to be similar to that of the West Midlands.

Conclusions

FISH and Q-PCR were both assessed in terms of the first two levels of the Fineberg evaluation hierarchy for diagnostic tests, that is whether the molecular test performs reliably and delivers precise information (technical capacity) and whether the test contributes to an accurate diagnosis of chromosome abnormalities (diagnostic accuracy).

When the technical capacity of FISH and Q-PCR is considered, a number of points emerge:

- The molecular tests currently available can be structured to either detect Down syndrome (trisomy 21) only, or the five most common chromosome abnormalities (trisomies 21, 13, 18 and X, Y sex chromosome abnormalities). No molecular test has been developed that can detect the whole range of abnormalities that karyotyping detects.
- In terms of detection of trisomy 21, both FISH and Q-PCR were judged to be as reliable and precise as karyotyping, where a conclusive test result could be obtained. Therefore, in terms of their technical capacity to test for Down syndrome, the study results indicate that both Q-PCR and FISH are potentially able to replace the current gold standard, karyotyping.
- In terms of the technical capacity to detect the five most common abnormalities (trisomies 21, 13, 18 and X, Y sex chromosome abnormalities), both molecular tests (i.e. commercial 5-probe FISH tests and

multiplex Q-PCR[11]) were also judged to be as reliable and precise as karyotyping, where a conclusive result could be obtained. Therefore, in terms of their technical capacity to test for the five most common chromosome abnormalities, the study results indicate that both molecular tests are potentially capable of replacing the current gold standard, karyotyping.

- In terms of test failure or inconclusive results, however, 5-probe FISH and multiplex Q-PCR exhibited a higher overall level of tests where no result was obtained (6.8% and 9.5% respectively) than did karyotyping (0.2%).
- Examination of the West Midlands and Northern Ireland series demonstrated that there were higher levels of inconclusive test results in Northern Ireland. Further analysis indicated that molecular methods appear to be sensitive to gestational age of the sample, with more likelihood of failure in early amniocentesis samples (11–13 weeks) and late samples (23 weeks plus). Maternal blood contamination may also lead to test failure and therefore the practice of excluding samples with visible bloodstaining should be considered. The extent to which molecular tests can be implemented routinely into service in the UK will be influenced by the proportion of contaminated samples normally received in local laboratories. The survey of UK laboratories reported in the next chapter collected information on this; laboratories reported that the percentage of bloodstained samples varied between 1% and 15%, with a mean of 5%.

Therefore, for amniotic fluid samples that are suitable for molecular testing, both FISH and Q-PCR demonstrated no evidence of significant differences in technical capacity (i.e. the molecular tests performed reliably and delivered precise information) when compared with karyotyping. However, early amniocentesis samples (11–13 weeks) and those contaminated with maternal blood may not be suitable for molecular testing.

The second level of the Fineberg hierarchy (i.e. diagnostic accuracy of the molecular tests) was assessed by identifying to what extent molecular tests failed to identify chromosome abnormalities that were detected by karyotyping. This was done in the following way:

- **Absolute sensitivity and specificity** were calculated based on the ability of the molecular test to detect all the abnormalities that the new test is capable of detecting compared to the gold standard (karyotyping) for these same abnormalities. This demonstrated high levels of absolute specificity for both types of molecular test (1.00 for FISH and 0.9997 for Q-PCR) and high absolute sensitivity (1.00 for FISH and 0.9565 for Q-PCR).
- **Relative sensitivity and specificity** values were also calculated for both tests, that is the ability to detect any abnormality within the range of possible abnormalities that can be detected by the gold standard (karyotyping). This calculation was based on technical performance as measured in the study and 11-year data from the laboratories on the numbers and types of abnormalities detected by karyotyping. Cases that would have been identified as high risk because of past medical history or family history were excluded, because they were likely to be offered full karyotyping, as were cases where the abnormality was judged not to be clinically significant. This demonstrated high relative specificities of 0.9999 for FISH and 0.9996 for Q-PCR. Relative sensitivities were also high, at 0.8605 and 0.8234 for 5-probe FISH and multiplex Q-PCR respectively; but were lower at 0.6478 and 0.6196 for the Down only FISH and Q-PCR tests.

[11] Triploidy can also be detected by the extended Q-PCR test.

Chapter 3

Current laboratory testing practices

Introduction

At the outset of the study, a number of NHS cytogenetics laboratories in the UK (especially the larger Regional Cytogenetics Laboratories) were already using FISH technology. This was generally as an adjunct to traditional cytogenetic tests. Anecdotally, clinicians reported they were requesting FISH as an add-on test, where available, to relieve anxiety caused by the longer wait for karyotyping results in cases where the risk factors were principally for Down syndrome. However, they were still relying on karyotyping as the definitive diagnostic test. Only one laboratory was known to be using FISH on interphase nuclei for routine prenatal diagnosis of Down syndrome (Lowther G, Duncan Guthrie Institute of Medical Genetics, Glasgow: personal communication, 1996).

In contrast, before the start of the study at the beginning of 1997, there had been no reports demonstrating the accuracy of large-scale studies of Q-PCR such as those reported for FISH. In the UK, few laboratories were using Q-PCR techniques for prenatal diagnosis. Furthermore, at the time of commissioning this study, the equipment required to analyse the fluorescent products of the Q-PCR test, a DNA sequencer, was generally only available in molecular genetics laboratories. Sequencers were therefore only accessible to cytogenetics laboratories on a limited basis, because molecular genetics laboratories are not necessarily located near the laboratories in which prenatal karyotyping is conducted.

As well as assessing the technical performance of molecular tests, the feasibility of introducing each molecular test into the routine test setting in UK cytogenetics laboratories was explored as part of this study. Other important factors that might influence the adoption of this new technology for use in prenatal diagnosis were investigated, for example availability of specialist equipment; the skills and competence of existing staff; established methods of working; and current and future organisation of laboratory services.

In 1991, a report by a working group of the Clinical Genetics Committee of the Royal College of Physicians[63] identified the need for more comprehensive genetic services organised on a regional level. The Committee argued that reorganisation would enable development (and retention) of expertise; maintenance of standards of investigations by internal and external quality control; and centralisation of equipment. The introduction of new genetic techniques might be an additional factor in the impetus to reconsider the geographical location of genetic laboratory services.*

UK survey of NHS genetics laboratories

In order to provide an accurate picture of existing cytogenetics services, a postal survey of all 35 genetics laboratories in the UK was despatched in May 1998. The survey questionnaire was targeted at the scientific director of each laboratory. Where such individuals were not responsible for prenatal tests, or for the possible introduction of molecular techniques, they were requested to pass the questionnaire on to the relevant person. Five laboratories reported that they did genetic testing but that this did not include prenatal samples; the remaining 30 returned a completed questionnaire. Twenty-nine laboratories (83%) reported that they currently used FISH, 48% on a daily basis, and 23 (66%) reported they used Q-PCR for some purpose.

Existing and planned use of molecular tests in NHS cytogenetics laboratories

Existing use of molecular tests in prenatal testing

Survey responses indicated that by mid-1998, seven of the 30 laboratories conducting prenatal testing already used FISH for prenatal diagnosis

* At the commencement of this HTA study in 1997, some NHS regions were served by a single genetics laboratory but several had two or more independently operated laboratories.

(two laboratories for all cases and five for selected cases). A smaller number (two laboratories) reported that in the index year 1997 they were using Q-PCR for prenatal diagnosis of Down syndrome, and one of these laboratories reported it was using both FISH and Q-PCR for the same purpose.

Of the 22 cytogenetics laboratories who were not using any molecular test for prenatal diagnosis, 12 reported that they were in the process of assessing FISH and three that they were assessing Q-PCR. No laboratory was assessing Q-PCR that was not already using or assessing FISH. Of the 12 laboratories that were assessing one or both molecular tests, only five gave a date for their planned introduction – all were during the year following the survey (i.e. in 1999). The remaining 10 laboratories (33%) reported that they were not yet considering the introduction and use of molecular methods for prenatal diagnosis. From responses it was clear that the potential use of molecular methods in routine prenatal testing was a question of great interest for the majority of laboratories, some reporting that they were waiting for the results of the HTA study.

No laboratory reported that it was using molecular tests as a replacement for karyotyping, that is where molecular tests were in use, even if they were used for all cases, the use of FISH or Q-PCR was in addition to karyotyping. Of the eight laboratories that were already using molecular tests for prenatal diagnosis in 1998, the majority had introduced the techniques within the previous 2 years (1996–98), although one laboratory reported having introduced FISH in 1992.

Anticipated use of molecular tests in future prenatal testing

Respondents were asked which of the two molecular tests they would prefer to introduce for the diagnosis of Down syndrome. *Table 13* shows that the largest number of respondents (13, or 47%) expressed a preference for FISH and 6 (20%) preferred Q-PCR. However, one-third of laboratories either expressed no preference or considered they lacked sufficient information to decide.

Some examples of the reasons given by respondents for their preference (or lack of preference) are also presented in *Table 13*. From these, it appears that FISH is preferred because the skills used in performing this molecular test are perceived as more closely akin to those used for karyotyping. Where Q-PCR is the preferred option, this is commonly on the grounds of perceived improvements in cost and efficiency. It should be noted that the mean throughput of laboratories preferring Q-PCR was double that of those stating a preference for FISH.

Respondents were also asked to anticipate how molecular test(s) might best be introduced into their prenatal diagnostic process, should they prove reliable. Of the laboratories that were considering introducing FISH (12 laboratories), 10 (83%) anticipated using commercially available kits rather than tests developed in-house, one laboratory was planning to use in-house tests and one had not decided. For Q-PCR, some laboratory directors identified the purchase of equipment as a significant barrier to the introduction of this test, although one respondent suggested that it would be easier for their service to identify funds for equipment than for retraining or appointing extra staff.

Only one respondent anticipated that molecular tests would in the future replace karyotyping; five laboratories (17%) foresaw molecular tests being carried out routinely in addition to karyotyping;

TABLE 13 *Geneticists' preferences for molecular tests used to diagnose Down syndrome*

Choices	Responses	Reasons
FISH	13	"We already have the necessary expertise and equipment for FISH…" "The technology fits in more easily with the existing laboratory set-up…" "We have greater experience of FISH techniques"
PCR	6	"More efficient, cheaper, more flexible…" "Batching of samples more effective for Q-PCR than FISH" "FISH is expensive, labour intensive…"
No strong preference	1	"…FISH and Q-PCR have associated problems"
Lack of information to choose	10	"Waiting for results of HTA"

seven respondents (23%) anticipated selective use of the tests. One laboratory wished to consult with users before decisions were taken. Sixteen respondents (53%) did not answer this question, that is approximately half of laboratories were unsure or had not considered this question.

Staffing and equipment for molecular tests in NHS cytogenetics laboratories

Existing staff and training needs

In order to explore some of the potential barriers to implementation of molecular tests for prenatal diagnosis, respondents were asked about current staff skills and anticipated training requirements. For both FISH and Q-PCR, laboratory directors (in laboratories not already using that particular molecular test for prenatal diagnosis) were asked about the staff training that might be required to successfully introduce each new technique into their laboratory; responses are summarised in *Table 14*.

For FISH, half the relevant laboratories (48%) reported that they already had enough skilled staff in post to be able to introduce the molecular test without further staff training, but for Q-PCR this was the case in only one laboratory. However, the majority of respondents (64%) were confident that they could train their staff 'in-house' if necessary for Q-PCR molecular testing, although this figure was even higher for FISH tests (96%). Moreover, there was no significant difference in the difficulty anticipated in recruiting skilled staff to conduct either molecular test, if this were necessary. Approximately two in three laboratories anticipated no problems with recruitment, regardless of which molecular test was introduced.

In order to understand better the nature of potential constraints on resources when introducing a particular molecular test, laboratory directors were asked whether they could afford to train staff in the new technique while continuing to provide a karyotyping service. *Table 14* illustrates that in 92% of laboratories the funding of training was perceived to be a major obstacle to the introduction of Q-PCR and in 74% if FISH were to be introduced. Finally, laboratories were asked for details of their staff and levels of seniority. The information on skill-mix was analysed across laboratories and used in the costing model developed for the study.

Existing equipment and its availability

As well as training staff in the new skills needed to undertake Q-PCR, introduction of this particular molecular technique also requires investment in (or access to) equipment not normally found in laboratories conducting prenatal diagnostic karyotyping. For Q-PCR, prepared samples are analysed, in batches, on a DNA sequencer (see page 9). The introduction of Q-PCR is therefore a capital-intensive option when compared to the introduction of FISH, but offers the advantage of batch processing, with the potential for lower staff input and employment of lower staff grades for preparatory work.

Only nine out of the 30 cytogenetics laboratories (30%) reported that they had access to a DNA sequencer with sufficient available capacity to allow it to be used routinely in prenatal testing. Five of these laboratories would have adequate access to one sequencer and four had access to two or three sequencers. A further three laboratories had very limited access to one or more sequencer(s). The remaining 18 laboratories reported no access to a DNA sequencer. There was no correlation between current access and number of samples processed per annum.

For the nine laboratories with access to sequencers, the types of DNA sequencer(s) available for prenatal testing are listed in *Table 15*. In all but two cases, sequencers were located within the same

TABLE 14 *Training needs if molecular tests introduced into service*

	FISH (%)	Q-PCR (%)
Total respondents	23	28
Enough skilled staff in post without further training?	11 (48)	1 (4)
Able to train 'in-house'?	22 (96)	18 (64)
Anticipate difficulties if recruiting necessary?	9 (39)	10 (36)
Sufficient funding to train staff in addition to carrying on with routine karyotyping if karyotyping replaced?	6 (26)	2 (7)

TABLE 15 *Types of DNA sequencers available for prenatal diagnosis*

Types of sequencers	No. of laboratories using model[a]	
	Access	Very limited access
ABI 373	2	1
ABI 377	2	1
ABI 310	3	1
ALF Express	2	

[a] *Some laboratories have access to more than one model*

department; the remaining two were accessible on-site, one on another floor and one 50 m away.

For the nine laboratories with sequencers:

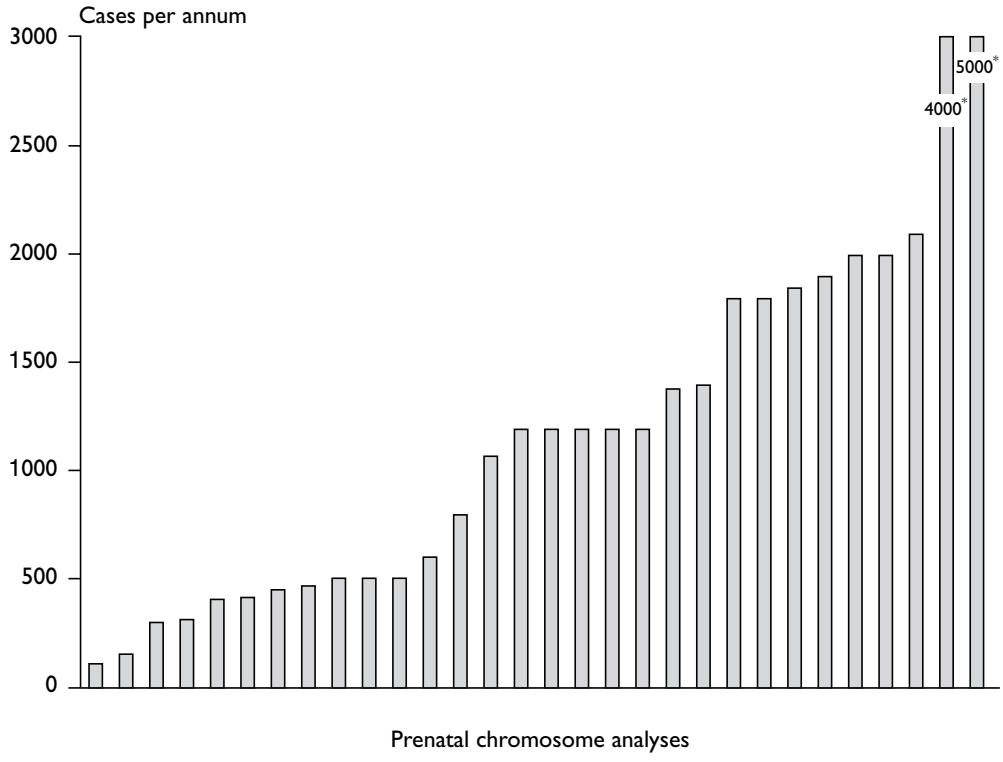
- Five laboratories predicted that use of the DNA sequencer for prenatal testing would not preclude access for other emerging applications of Q-PCR.
- One laboratory stated that the level of use of its sequencer, currently heavily used for detection of single gene defects, might be a barrier to full introduction of Q-PCR for prenatal testing using this equipment.
- One laboratory reported that it already used the sequencer 24 hours per day over 7 days to fulfil R&D commitments, which would severely limit its service use.

Prenatal services and activity in NHS cytogenetics laboratories

Volume of prenatal work and hours of service

In addition to asking questions about staffing and equipment, the survey questionnaire also explored current activity levels for prenatal testing. All 30 laboratories undertaking prenatal testing provided information on their service activity. In the index year, 37,780 amniotic fluid samples were reported to be processed by these 30 laboratories, with the smallest laboratory processing approximately 100 samples and the largest approximately 5000 (*Figure 2*).

The reported UK mean (and median) throughput was 1200 samples, with the 25th and 75th percentiles being 500 and 1800 samples per annum,



* *Shows two extreme values that have been curtailed on the graph*

FIGURE 2 *Prenatal chromosome analyses conducted in UK laboratories*

respectively (rounded to the nearest 50 samples). For ease of interpretation of results a mean throughput of 1000 samples per annum has been used for calculations in this report, as 47% of laboratories reported processing 1000 or fewer amniotic fluid samples for karyotyping per annum. The reported mean size of amniotic fluid samples received by these laboratories is in the range 6–20 ml (median 15 ml).

In the 25 laboratories that reported handling chorionic villus samples as a part of their workload, a total of approximately 4500 CVS specimens were reported as received for the index year (range 16–580, median 130). The ratio of chorionic villus samples to amniotic fluid samples received in these laboratories was approximately one in five (median 11%, range 4–88%). All but one of the laboratories were able to report the percentage of samples that were 'bloodstained'; this ranged from 1% to 15% of all samples (mean 5%). Twenty-six laboratories out of 30 reported a total of 759 Down syndrome cases detected in 1997; the mean prevalence of those tested was calculated to be 2.1% (range 0–5.3%). Similarly, not all laboratories were able to report the number and types of other chromosome abnormalities detected by prenatal testing. Twenty-three laboratories were able to report the number of X, Y chromosome gender abnormalities, with 39 XXX and 39 XXY detected in 1997 (mean prevalence 0.2%). Finally, 777 other abnormalities (including trisomies 13 and 18) were reported from 23 laboratories (mean prevalence 2.9%). These figures are similar to those reported in the two study laboratories over an extended period.

Reporting times for test results and hours of working

Laboratory directors were asked for the average turnaround time for karyotyping between receipt of sample and despatch of report. The range of reporting times is shown in *Figure 3,* with a mean time of 13 days (range 7–17 days).

In a separate survey of UK obstetricians (reported later, in *Figure 13*), clinicians were also asked what patients were told about when results might be expected (median 15–21 days). The reporting times that obstetricians advised were significantly longer than the mean turnaround time quoted by clinical scientists and were closer to the maximum of the range reported from the slowest laboratories.

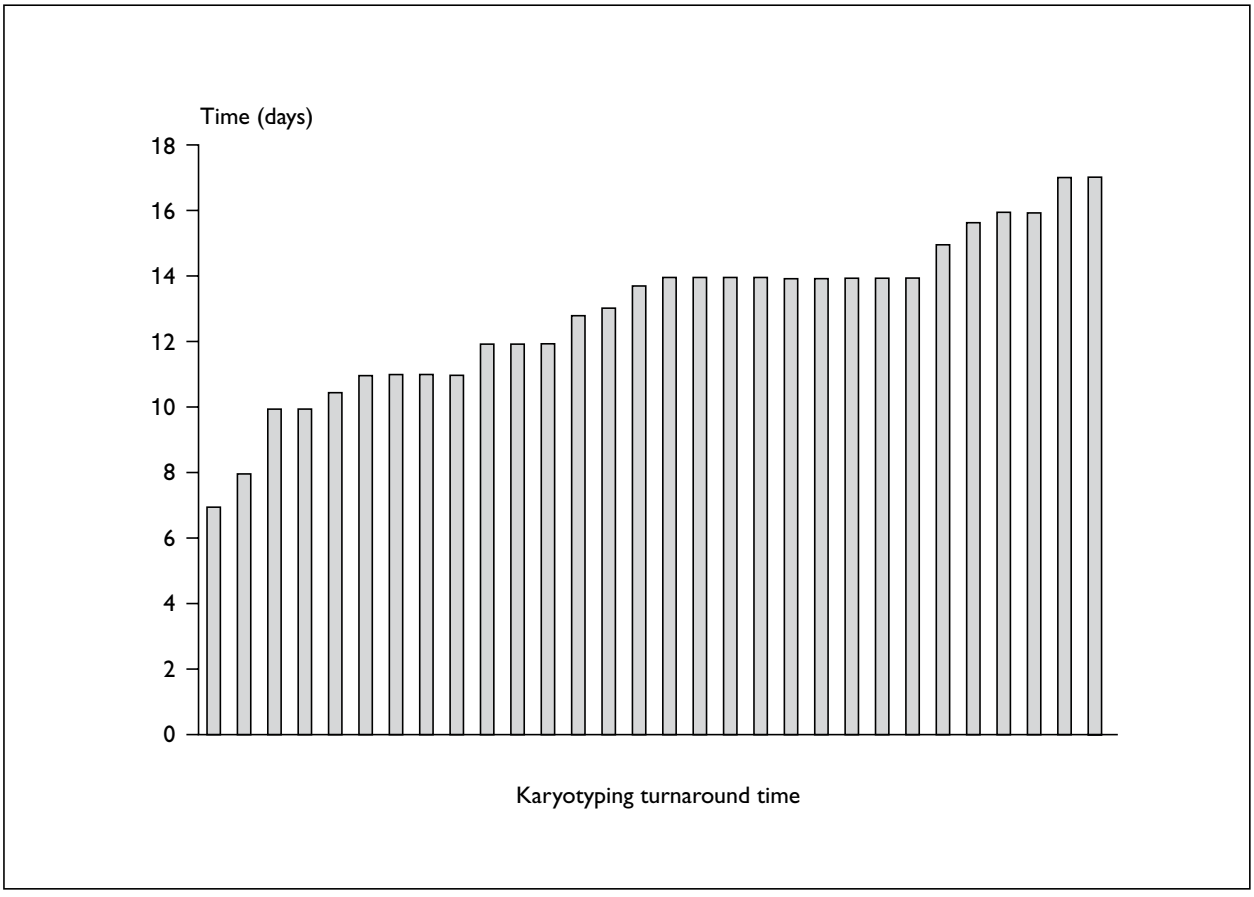


FIGURE 3 *Karyotyping turnaround time reported by clinical scientists*

If laboratories are to expand their workload to include molecular testing as an adjunct to karyotyping, an extension of hours of working and/or an expansion of staff and premises might be required. All 30 laboratories reported on their hours of work. Mean working hours during the week were 9 hours per day (range 7–13 hours). Seventeen laboratories (57%) worked on a Saturday for a mean of 4 hours (range 1–10 hours), and ten laboratories worked for a mean of 4 hours on a Sunday (range 1–7 hours).

Study sites: structuring of prenatal testing services

Although laboratory and clinical activity with respect to prenatal testing for chromosome abnormalities has been described nationally up to this stage, a detailed assessment of molecular tests was primarily carried out in two areas of the UK. One area, the West Midlands, has a cytogenetics laboratory with one of the largest throughputs in the UK, and the other area, Northern Ireland, has one of the smallest. The organisation of services for prenatal chromosome analysis in both communities (and thus for supplying cases to the study) was somewhat different. These differences are described below.

The West Midlands

Figure 4 illustrates routes for access to prenatal chromosome testing in the West Midlands. The clinical genetics laboratory and tertiary services for the whole of the region are based at the Women's Hospital Trust, Birmingham.

For routine prenatal testing, each hospital in the region has a serum screening programme. It has been agreed at a regional level that the accepted threshold risk for requesting amniocentesis should be 1 in 280 with respect to Down syndrome.

In the West Midlands, all counselling prior to screening or amniocentesis is carried out locally. In most Trusts a midwife with expertise in counselling is involved in prenatal testing. In some cases this responsibility is spread across a team of midwives, who work in the community as well as within the acute Trust. All antenatal clinics can call upon genetics associates via the Regional Genetic Laboratory. Very high-risk cases, or those with a recognised risk of single gene defects, are usually referred to the tertiary services where genetic counselling can be provided before testing. All samples for testing are forwarded to the Regional Genetic Laboratory, with each Trust making its own arrangements for transport of samples.

Within the laboratory, all samples are processed on the day of receipt, and cultures first examined after 5 or 6 days. Cultures are harvested when sufficient cell growth is available for karyotyping (usually between 8 and 12 days from receipt of

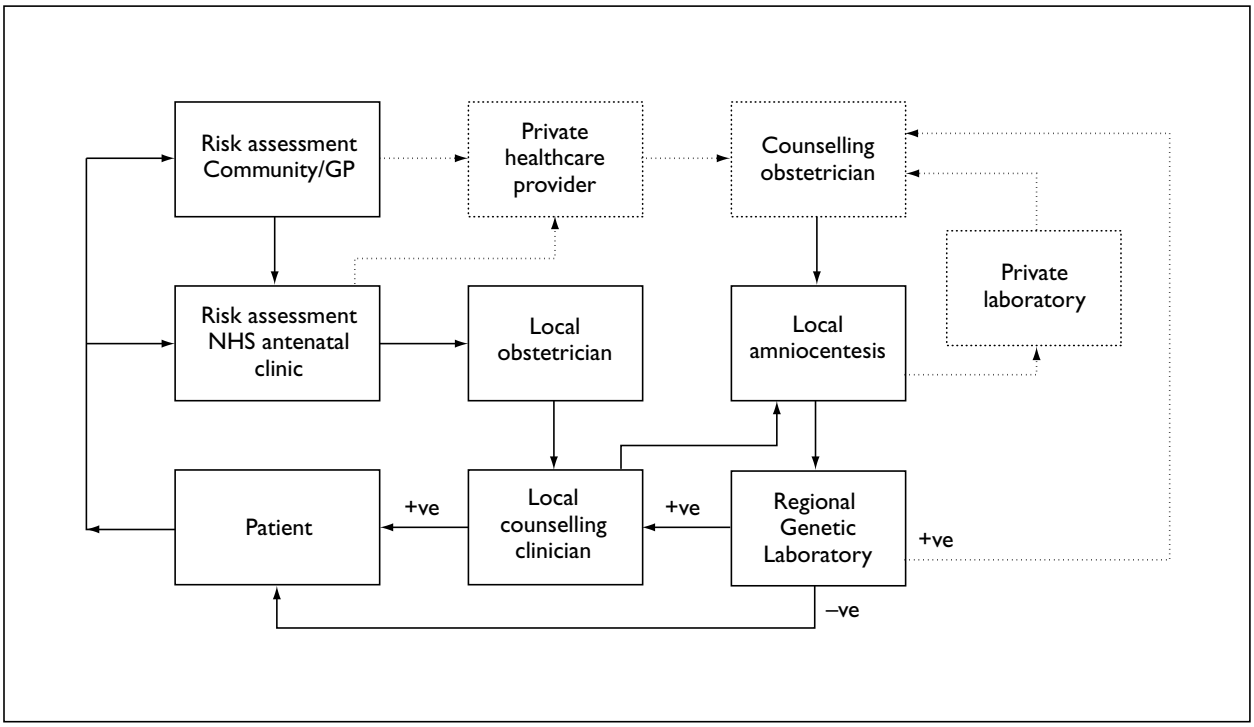


FIGURE 4 *Routes for access to prenatal chromosome testing in the West Midlands*

sample). A detailed analysis of the chromosomes is carried out and checked by an independent scientist. Results are mailed by the laboratory direct to the patient, unless the result is abnormal. In abnormal cases, the scientist phones the referring clinician (obstetrician or midwife, as available) with the result.

Northern Ireland

The routes for access to prenatal chromosome testing in Northern Ireland are shown in *Figure 5*. Prenatal diagnosis for the whole of Northern Ireland is organised centrally through the Northern Ireland Regional Genetics Centre, based at Belfast City Hospital. Obstetricians, GPs and patients themselves can contact the department and request an appointment at one of the prenatal diagnosis clinics. These clinics are held weekly in the Royal Maternity and Jubilee Maternity hospitals in Belfast, and fortnightly in Craigavon Area Hospital and Antrim Area Hospital. At least one obstetrician and one clinical geneticist is present at each clinic. Patients are allocated to these clinics according to the reason for the referral and geographical convenience.

All the clinics can perform ultrasound scans and amniocentesis.[†] In practice, many of those referred with abnormal scans are referred to the Royal Maternity Hospital Belfast. All patients attending these clinics are seen initially by a clinical geneticist for counselling and risk assessment. They are then seen by the obstetrician at the clinic for their scan and any procedures, such as amniocentesis. If the scan is abnormal, patients will probably be counselled again by the geneticist.

All amniocentesis samples are analysed by the Northern Ireland Regional Genetics Centre, based at Belfast City Hospital. Patients are able to phone the laboratory directly for the result and a written copy of the result is then sent to the referring

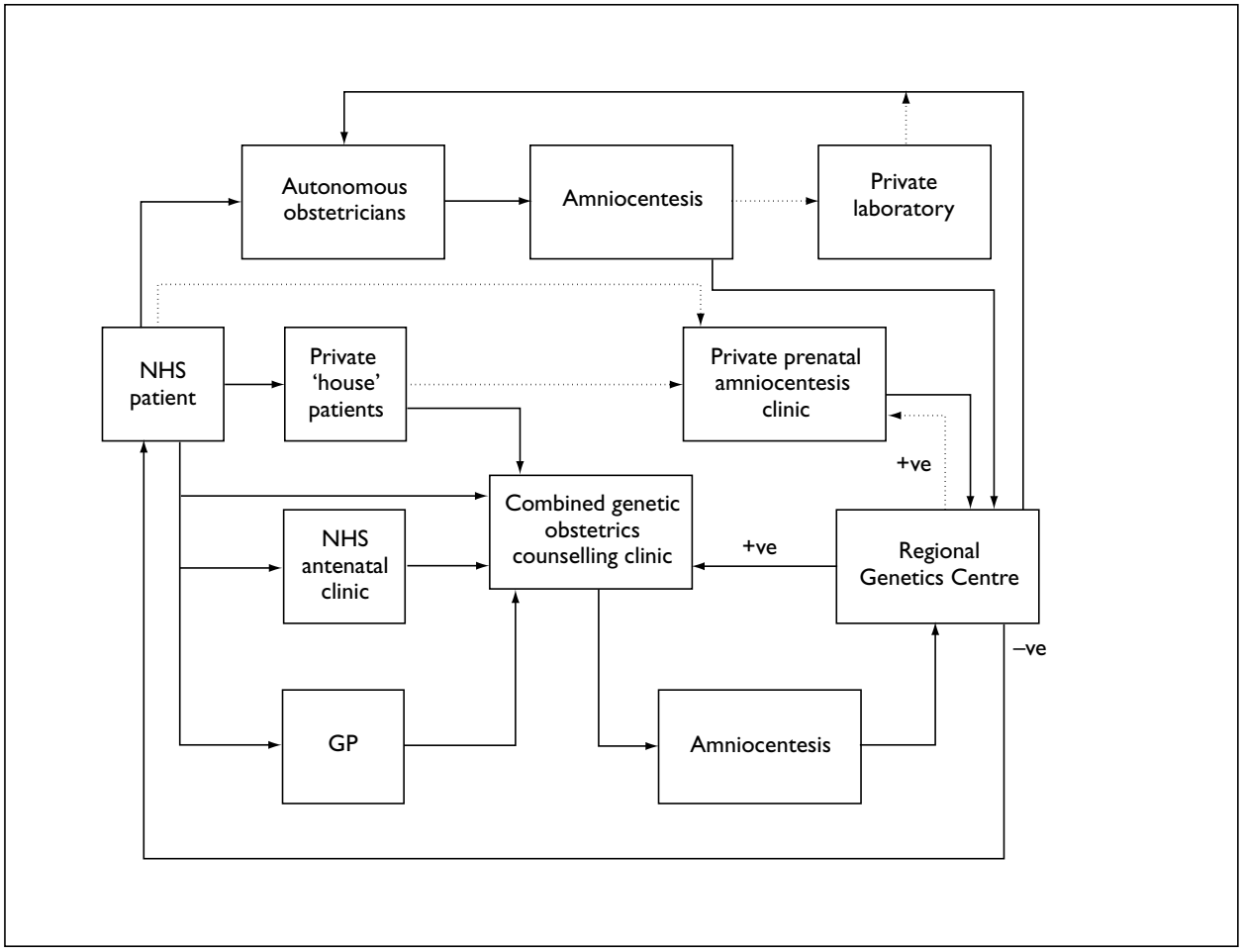


FIGURE 5 *Routes for access to prenatal chromosome testing in Northern Ireland*

[†] Chorionic villus biopsy can be done at both Belfast clinics and cordocentesis can be carried out at the Royal Maternity Hospital.

doctor. Abnormal results are usually conveyed by one of the clinicians.

There are, however, some exceptions to the process described above. The Ulster Hospital NHS Trust organises its own prenatal diagnosis clinics, with no input from the clinical geneticists. Samples are sent directly to the prenatal diagnosis laboratory of the Northern Ireland Regional Genetics Centre at Belfast City Hospital and results are conveyed directly to the obstetrician concerned. One consultant at the Royal Maternity performs his own amniocentesis tests with no input from the clinical geneticists and a number of patients are seen in the private prenatal diagnosis clinic. However, the majority of patients having prenatal diagnosis in Northern Ireland are seen by a clinical geneticist within the Northern Ireland Regional Genetics Centre.

Conclusions

Information obtained from laboratories undertaking prenatal chromosome analysis supports the assumption made in chapter 2 (*Figure 1*) that the adoption of FISH is more advanced than Q-PCR. It was evident from the survey that the reasons for this are two-fold. Firstly, the skills needed to perform FISH are similar to those required for karyotyping and secondly, a much larger cultural shift and higher training input is required to adopt Q-PCR. Nevertheless, laboratory directors were confident this could be achieved within the current service. However, in the free text components of the survey many laboratory directors expressed reservations about the completeness of the research evidence on the reliability of molecular tests as a replacement for karyotyping.

In terms of staffing and equipment, it would appear that current cytogenetic staff are implementing, or will be able to implement, both forms of molecular test, although some laboratories noted general graduate recruitment problems.

In terms of the likely cost of introducing particular molecular tests in UK laboratories, Q-PCR poses the greater challenge. The DNA sequencer required to implement Q-PCR testing would be the most expensive single item of equipment within a cytogenetics laboratory, and there is no evidence of spare capacity on the few machines currently accessed for ad hoc prenatal chromosome analysis. However, in relative terms, the cost of a sequencer is comparable with the cost of single ultrasound machines purchased elsewhere in antenatal services.

The size of samples received by laboratories may also be a constraining factor. One possible model for the diffusion of molecular tests is their introduction in addition to karyotyping, that is both a molecular test and karyotyping performed for all patients. It has been suggested that a minimum sample size of 15 ml would be required to routinely perform both a molecular test and karyotyping. The survey responses confirm that there is wide variation in the size of samples received by cytogenetics laboratories. These differences arise from natural variations between cases in the amount of amniotic fluid available to aspirate, the proportion of first trimester amniocenteses conducted in feeder antenatal services, and the expertise of clinicians in drawing samples. It is clear that it would not be possible to perform both tests in all cases, and this may have to be decided once a sample has reached the laboratory. These circumstances may place additional burdens on counselling.

Finally, general comments from laboratory directors confirmed earlier discussions in this report, in particular on page 17 ('Clinical significance'). Respondents placed particular emphasis on the difference between a molecular test, which can be used to exclude specific abnormalities, and karyotyping, where comprehensive assessment of abnormalities is possible. If molecular tests were to replace karyotyping, it is argued that this would require a major shift in the way in which prenatal testing is viewed.

Chapter 4

Costs of karyotyping and molecular tests

Introduction

In order to assess the cost-effectiveness of introducing molecular tests for prenatal diagnosis of chromosome abnormalities, the study had first to estimate the cost of performing the various tests, and then compare the resources used for a number of different testing scenarios. A combination of a primary costing approach and cost modelling was used to estimate costs for various routine service settings. A detailed cost analysis was carried out first for the three main diagnostic tests studied – karyotyping, FISH and Q-PCR. This was conducted from the perspective of the NHS, and the principal focus was on the direct cost of performing each diagnostic test. Costs not associated with diagnostic test performance (such as those of prenatal test counselling, amniocentesis, sample transport, laboratory reception, despatch of test results, and other routine laboratory clerical work) were expected to be independent of the type of diagnostic test, and were therefore excluded from the cost analysis.

An initial overview of laboratory operations was used to identify the main resource differences between the three tests. Major differences were identified in the use of capital equipment, staffing levels and associated skill-mix, and consumables. Direct observation and measurement of work patterns then enabled quantification of resource use for each test separately. This information was supplemented by routine data on activity and other data from the national survey of laboratories. In this way, estimates based on direct observation were supplemented by an indication of UK ranges, where possible. The observational cost study was principally conducted in the West Midlands Regional Genetic Laboratory and Consultancy Services, Birmingham, with some cost data also being provided by the Northern Ireland Regional Genetics Centre, Belfast.

Resource use was measured for each test using the following three main categories:

- **Capital** – equipment regularly used was identified and valued based on purchase price expressed as an equivalent annual cost, using standard discounting procedures. Accommodation costs were based on the annual charge levied by Trusts plus a formula devised to estimate the space occupancy associated with each test. Overhead costs such as heating, lighting, cleaning, etc. were incorporated into accommodation costs.
- **Labour** – predominantly laboratory staff time, this was measured as the amount of time, on average, spent in dealing with samples, and valued by the wage rate per hour for the grade of staff performing the procedure.*
- **Consumables** – items such as test kits, reagents and disposable items were quantified and valued at market rates.

The first category (capital) consists of items that are essentially fixed costs, in that they occur irrespective of the level of laboratory activity. The second category (staff cost) is semi-fixed in the medium term, although it can be considered variable in the long run. The final category, consumables, represents a true variable cost, incurred only when a test is performed.

In order to value the cost of services, it was assumed that average costs reasonably reflect long run costs. The marginal cost consequences of different tests, and variations arising from alternative patterns of testing, were modelled based on this assumption. In addition, the costing study sought to identify and estimate any economies of scale, and to examine break-even points under particular laboratory conditions. All costs were based on 1998 prices.

The following sections outline the approach used to derive the costs of the two main types of molecular test (tests for Down syndrome only and 5-probe/multiplex tests) and karyotyping. A separate sensitivity analysis section is not presented, instead variations arising from changes to inputs in the modelling process are discussed in the relevant sections.

* Including Clinical Scientist, Medical Technical Officer (MTO) or Medical Laboratory Assistant (MLA).

Capital costs

The main capital costs associated with prenatal testing are equipment and accommodation costs.

Equipment costs

All items of equipment used for each test were first identified and then an equivalent annual cost calculated to include depreciation based on market price and the Treasury discount rate (of 6%). All equipment was assumed to have a life of 5 years; this assumes that technological advances will necessitate purchase of new equipment before the original machines become unserviceable. At the rate of advance of technology in this area, a 5-year life is probably a conservative estimate.

The capital equipment used in a cytogenetics laboratory may fall into one of two main categories: (i) dedicated equipment used exclusively for prenatal chromosome diagnosis or (ii) general genetic laboratory equipment. In the laboratory in which the main costing study was carried out, laboratory equipment that might be used in karyotyping, FISH or Q-PCR was primarily dedicated to prenatal testing for most of the time. The assumption was made in this analysis that such equipment was dedicated to prenatal testing and, as a result, the capital equipment cost calculated may represent an overestimate for smaller laboratories in which some sharing of equipment takes place between prenatal diagnosis and other work of the genetics service. However, any reduction in the capital costs allocated to prenatal testing in such laboratories is likely to be counterbalanced by a lower prenatal test throughput, resulting in a similar cost per test performed. The relationship between fixed costs and throughput is explored more fully on page 37 ('Fixed costs').

Allowance was also made for the maintenance costs of any specialist capital equipment required for testing. Maintenance contract costs were most relevant in the case of Q-PCR, which is dependent on the use of a complex piece of machinery, the DNA sequencer. The other two tests, FISH and karyotyping, both use equipment that is less technologically complex, more robust, and therefore less likely to fail or require a rapid service maintenance contract. Of the nine UK laboratories that reported, in the national survey, having access to a DNA sequencer, only a small number had access to more than one sequencer; it would therefore appear that most cytogenetics laboratories will not have access to a back-up analyser in case of equipment malfunction. As a result, if the Q-PCR test were to be offered in a service setting it is likely that a maintenance contract, structured for rapid response, would be essential. Costs for such a maintenance contract were therefore included in the cost of Q-PCR testing, averaged over the 5-year DNA sequencer equipment life (allowing for an initial 12-month guarantee period).

Items of office equipment were not included in equipment cost estimates because these are generally used for tasks such as typing out reports and booking in samples. These activities are common to whichever test is used and, as described on page 33, were therefore excluded from the costing exercise.

Accommodation costs

In the cost analysis it was assumed that molecular testing would continue to take place in the same laboratory accommodation. However, different tests might require different amounts of space, particularly linked to the degree to which a test is labour-intensive as opposed to equipment-intensive.

Accommodation costs were calculated for each test, based on the annual charge levied by the Trust. This charge was levied over the genetic service as a whole, and not just for the prenatal testing component of the service. Therefore, in order to approximate accommodation costs for each individual prenatal test, a simple distributive approach was used to estimate a 'notional' accommodation cost. The total departmental accommodation cost was first distributed across the number of scientific staff (whole time equivalents) involved in testing.[†] Accommodation costs were then allocated based on the staff time dedicated to each type of test; it was assumed that the floor space required for capital equipment was relatively small in comparison to that required for staff and could therefore be ignored. Although an approximation, this method enabled an estimation to be made of the amount of laboratory space occupied by staff for particular test performance, and allowed a charge to be placed on this. This approximation was judged to be the most accurate for calculation of accommodation costs where a mixture of molecular methods and karyotyping may be conducted side by side in the same laboratory space. Thus, accommodation costs calculated in this way

[†] In this exercise, the staffing of a large laboratory was used for these calculations.

TABLE 16 *Equipment and accommodation costs for the three test methods*

	FISH	Q-PCR	Karyotyping
Annual cost: equipment[a] and accommodation	£4,230	£25,526	£3,400
Maintenance contract	Not available	£3,504	Not available
Total annual capital cost	£4,230	£29,030	£3,400

[a] *Assumes equipment life of 5 years*

will be sensitive to the staffing levels of laboratories, which, in turn, will be linked to throughput. Labour costs are discussed in the next section.

Table 16 lists annual equipment, accommodation and maintenance costs for the three test methods. These total annual capital cost figures are based on a UK mean laboratory throughput of 1000 cases per annum.

Labour costs

Annual labour costs will vary depending on the type of test being carried out, and the number and mix of staff employed to perform these tests. As pointed out previously, although staff costs are fixed in the short term, in the longer term they can be considered variable. In the present study, staffing costs were therefore treated, taking the longer time horizon, as variable costs. It was also assumed that staff could shift relatively easily between test methods at a defined skill level, although respondents to the laboratory survey did indicate that some retraining might be necessary.

To calculate annual labour costs, each testing process was first observed and broken down into a series of discrete tasks. Next, the direct staff input to each particular task was measured through observation in the service laboratory setting of the time required for completion, together with the level of staff performing the task.

In order to identify the appropriate grade of staff required for each discrete task, expert opinion was next used to define the actual level of skill required for each part of the testing process. Once the time taken for completion of each part of the test had been assessed within the laboratory setting, overall labour costs were calculated based on the grade of scientist required to conduct each procedure. For all staff grades, the mean of the appropriate pay scale was used to estimate staff costs. The broad task areas identified and the associated skill levels are shown in *Table 17*.

TABLE 17 *Grade of staff required to perform the components of each testing process*

Test	Process	Staff grade
FISH	Slide preparation and hybridisation	MLA
Q-PCR	Extraction of DNA	MLA
	Gel preparation	MLA
	Amplification and sequencer set-up	MTO1
FISH and Q-PCR	Analysis	MTO1
Karyotyping	Setting up	MLA
	First look	MTO2
	Changing culture medium	MLA
	Harvest	MLA
	Preparing slides	MLA
	Staining/banding	MLA
	Analysis	MTO2
All	Reporting results	Senior grade[a]

[a] *'Grade' may vary depending on level of seniority required for reporting in a laboratory*

The staff cost calculated for each test will also vary to some extent, depending on local practice and laboratory efficiency. For example, for some components of the testing procedure, a batching process was observed to increase throughput and reduce the staff time required per sample. These variations were incorporated into the final cost modelling.

Staffing levels (and grades) in a particular laboratory might be a function of the history of the laboratory as well as the actual task requirements. Thus, in many laboratories there may be a divergence between the optimum skill-mix and actual staffing levels. The laboratory survey demonstrated significant variations in skill-mix between laboratories.

Some of this skill-mix difference can be attributed to natural differences linked to laboratory size. In a small laboratory that processes fewer prenatal specimens, more senior staff may have to perform tasks that would otherwise be delegated to a junior member of staff in a larger laboratory, because their time is not fully occupied with higher level tasks. The approach adopted for measuring staff costs in the present study may therefore result in a slight underestimation of staff costs in smaller laboratories.

The relationship between staffing levels, skill-mix, batching and throughput is complex and could only be approximated in the cost model. Batching of samples, particularly for Q-PCR, does not appear to have a major effect in absolute terms on labour costs. For example, although the most common DNA sequencer is capable of carrying up to 36 reactions[‡] in one sequencer run, the sample preparation stage remains the limiting factor in the process. Any actual staff cost saving resulting from such large-scale batching will be limited to the relatively small amount of loading time saved. Similarly, the range of abnormalities to be detected (Down syndrome only or trisomies 21, 13, 18 and numerical sex chromosome abnormalities) also does not appear to significantly affect staffing costs. Long run staff costs are principally influenced by throughput and skill-mix; this is considered in greater detail on page 41 ('Trends in laboratory organisation (and skill-mix)').

Consumable costs

The consumable costs associated with all three tests include test kits, reagents and various small laboratory items. These costs were once again estimated through observation of laboratory practice. Because consumable costs are related to the number of samples processed, they represent true variable costs and annual costs can therefore be estimated directly in terms of a cost per sample tested and throughput.

Test kits, reagents and other consumables

The main consumable cost for FISH was the cost of the commercial testing kit. A single-probe FISH kit for trisomy 21 is less expensive than a 5-probe kit for detecting trisomies 21, 13, 18 and numerical sex chromosome abnormalities. When a commercial FISH test kit is used,[§] this component of the cost per sample tested remains constant. However, other consumable costs (such as reagents) may vary, depending on the extent of batching during earlier processing of samples. In the laboratories observed, batch size was generally small for FISH tests and the influence of batch size was therefore also relatively small.

For Q-PCR, which used in-house materials rather than commercial kits, consumable costs were generally far lower than for FISH tests. Apart from the cost of primer and amplification medium, the main Q-PCR consumable costs were those associated with preparing samples for the DNA sequencer. Because sequencer reactions may be run in very large batches, batch size was observed to influence consumable costs more significantly for this test. For example, a gel prepared for use in a DNA analyser might be used for a single specimen or for 24 reactions.

Small laboratory items

Other costs included as consumables were small laboratory items such as slides, pipettes, pipette

[‡] In practice some of these lanes are always allocated to controls and for maximum efficiency not all lanes will be used for reactions for prenatal diagnosis. For this analysis, it has been assumed that 20 lanes will be used per sequencer run.

[§] In the national survey, all laboratories using FISH for prenatal chromosome analysis in 1997 reported that they used commercial kits. One laboratory that was assessing FISH reported using in-house tests, and one laboratory not yet using FISH reported that they had plans to develop in-house tests.

tips, etc. In some cases the items used were reusable, while other items were disposable. Reliable information concerning the life expectancy of reusable items, and the cost of cleaning them, was not easily determinable. Thus, all such small laboratory items were costed as disposable items.

Royalty payment

In addition to the consumables used in testing, the new tests may incur a royalty payment. FISH does not have royalties associated with its use and, in the research setting, the use of Q-PCR did not carry a royalty payment. However, once in routine use Q-PCR will incur a royalty payment. This has been negotiated by the NHS at £2.70 per test for a single marker, and £5.40 per test for multiple markers (e.g. multiplex Q-PCR). The NHS will incur these costs until 2006, the year of expiry of the patent.

The total annual cost of consumables per sample tested (i.e. test kits, reagents, small laboratory items, royalty payments, etc.) is presented in *Table 18* for each molecular test (Down only and 5-probe or multiplex form of test) and for karyotyping. This table (based on a throughput of 1000 samples) also shows what proportion of the consumables is sensitive to cost reductions by batching when processing samples.

The estimated proportion of costs sensitive to batching is based on the maximum batch size possible for each test. Q-PCR demonstrates higher efficiency gains because more of the processes can be conducted with larger batch sizes. Consumable costs will rise, however, if optimum batch sizes are rarely achieved, because fewer specimens are received per day (low throughput laboratories) or the timing of arrival of specimens during the day precludes optimum batching. The influence of batch size is discussed further below.

Cost per sample tested

The total cost per sample (i.e. capital, labour and consumables) was estimated for laboratories with different throughputs. The results of the survey of UK laboratories, illustrated in *Figure 13,* had demonstrated a considerable variation in the number of prenatal chromosome analyses conducted each year.

Fixed costs

Table 19 shows the estimated total fixed costs (equipment and accommodation) per sample tested for FISH, Q-PCR and karyotyping at various laboratory throughputs. The annual fixed cost for each test has been converted to a fixed cost per test by dividing it by laboratory annual throughput in the range 100 to 5000. The fixed cost calculation assumes that one DNA analyser would be sufficient for all throughputs; this is reasonable because the maximum throughput for an analyser working 7 days per week is estimated to be 28,000 reactions per annum.

Variable costs

The two main variable costs are consumables (reagents and kits) and long run staff costs.

The consumable cost per sample tested is influenced principally by the type of test being performed. However, it may also be affected by (i) the range of abnormalities to be tested for and (ii) the extent of batching during processing, as discussed on page 36 ('Consumable costs').

TABLE 18 *Annual consumable costs and influence of batching*

Test	Proportion of total per sample tested (%)		Total annual consumable cost[a]
	Sensitive to batching	Not sensitive to batching	
FISH Down syndrome only	12.9[b]	87.1	£23,560
FISH 5-probe test	5.5[b]	94.5	£56,500
Q-PCR Down syndrome and multiplex	28.9[c]	71.1	£9,440
Karyotyping	15.9[d]	84.1	£11,520

[a] *Calculated for a throughput of 1000 samples per annum*
[b] *FISH costs calculated for preparation batch size of eight per batch*
[c] *For Q-PCR all calculations are based on 20 lanes used per DNA sequencer run*
[d] *For karyotyping it is assumed samples are processed in batches of four*

TABLE 19 *The influence of throughput on fixed cost per test*

Throughput (samples per annum)	UK laboratories with this volume (cumulative %)	Cost per test (equipment and accommodation) (£)		
		FISH Down and 5-probe	Q-PCR Down and multiplex test[a]	Karyotyping
100	3.3	50.35	293.31	58.64
500	25	16.51	61.08	31.44
1000	50	12.28	32.05	28.04
1500	75	10.87	22.37	26.91
5000	100	8.90	8.82	25.32

[a] *Excluding royalty payment*

Currently, the two main options for molecular tests, in terms of the range of abnormalities tested for, are (i) to test for Down syndrome only, or (ii) to test for the five most common abnormalities (trisomies 21, 18 and 13 and X, Y chromosome abnormalities). *Table 20* shows that, for FISH, an increase in the range of abnormalities to be detected produces relatively large differences in estimated variable costs, arising largely from increased consumable costs. Commercially available kits were used for FISH tests in this study, although one laboratory in the UK was trialling in-house FISH. However, an estimate of costs from this source was not available at the time of the study. For Q-PCR, as can be seen from *Table 18,* the costs of consumables (excluding royalties) are regarded as virtually equivalent, because there is only a difference of a few pence in the cost of consumables between the Down only test and a multiplex Q-PCR test. For Q-PCR, expert opinion suggests that in-house tests, rather than commercial kits, will be used in any roll-out of the technique into prenatal testing.

The variable cost components for Q-PCR, including labour costs, were determined at the two centres involved in this study, which were both using in-house tests. *Table 20* shows that Q-PCR is relatively more sensitive to variations in batch size. The figures for Q-PCR reveal that it is possible to reduce variable costs by maximising the batch size within practical limits set by the nature of the processes. This is because the initial cost of consumables for Q-PCR is low, so a moderate change in cost with changing batch size can produce a substantial percentage change in the overall cost per test. For laboratories that are receiving few amniotic fluid samples, or are receiving variable quantities of samples throughout the week, batch sizes that optimise the process may not be possible and this will increase the variable cost per test when averaged over the year.

Finally, labour costs per test are influenced principally by the skill-mix in the laboratory in which the molecular tests are carried out. The laboratory in which costs for this study were measured was large and able to incorporate an optimal skill-mix pattern (*Table 17*). This issue is discussed further on page 41 ('Trends in laboratory organisation (and skill-mix)').

Total cost breakdown for different tests

Figure 6 provides a breakdown of the total cost per sample tested for FISH, Q-PCR and

TABLE 20 *The influence of batch size and test regime on variable costs (at laboratory median throughput 1000 per annum)*

Number of samples per batch	Variable costs per test, £			
	8	6 (%[a])	4 (%[a])	1 (%[a])
FISH (Down syndrome only)	29.86	30.25 (1.3)	31.03 (3.9)	38.07 (27.5)
FISH (5-probe)	62.79	63.05 (0.4)	63.58 (1.3)	68.31 (8.8)
Q-PCR[b]	12.11	12.59 (4.0)	13.56 (11.9)	22.22 (83.5)
Karyotyping			61.64 (0.0)	64.07 (3.9)

[a] *Percentage change from optimum batch size, assumed to be eight for molecular tests and four for karyotyping*
[b] *Percentage difference in consumables between Q-PCR Down syndrome and multiplex is negligible*

karyotyping. Costs are broken down into fixed (equipment and accommodation) and variable (consumables and staff). This figure demonstrates that for the three tests the most significant cost elements are:

- FISH – the major element is the cost of consumables (i.e. test kits).
- Q-PCR – the major cost is associated with the capital cost of the equipment required (i.e. DNA analyser).
- Karyotyping – the largest element of the cost of a test is the cost of labour.

Total cost per test for different throughputs

Table 21 shows the impact that changes in throughput have on the calculated total cost per test.[¶] Both molecular tests and karyotyping naturally show a cost reduction with increasing throughput. This effect is most noticeable for Q-PCR because of the capital-intensive nature of this investigation (*Figure 6*) and the low level of variable costs. FISH and karyotyping costs are far less sensitive to throughput.

The two molecular tests also differ in the sensitivity of their costs to the range of abnormalities to be detected. For Q-PCR, a move from a Down syndrome only test to a multiplex test at a throughput of 1000 cases per annum leads to only a small increase (7.4%) in the cost per test. For FISH, however, the cost per test changes significantly; a 5-probe test is 78% more expensive than Down syndrome only at a throughput of 1000 samples per annum.

Figure 7 illustrates the relationship between throughput and cost per test. In this figure, the vertical lines represent the percentage of UK laboratories processing this number of samples per annum.

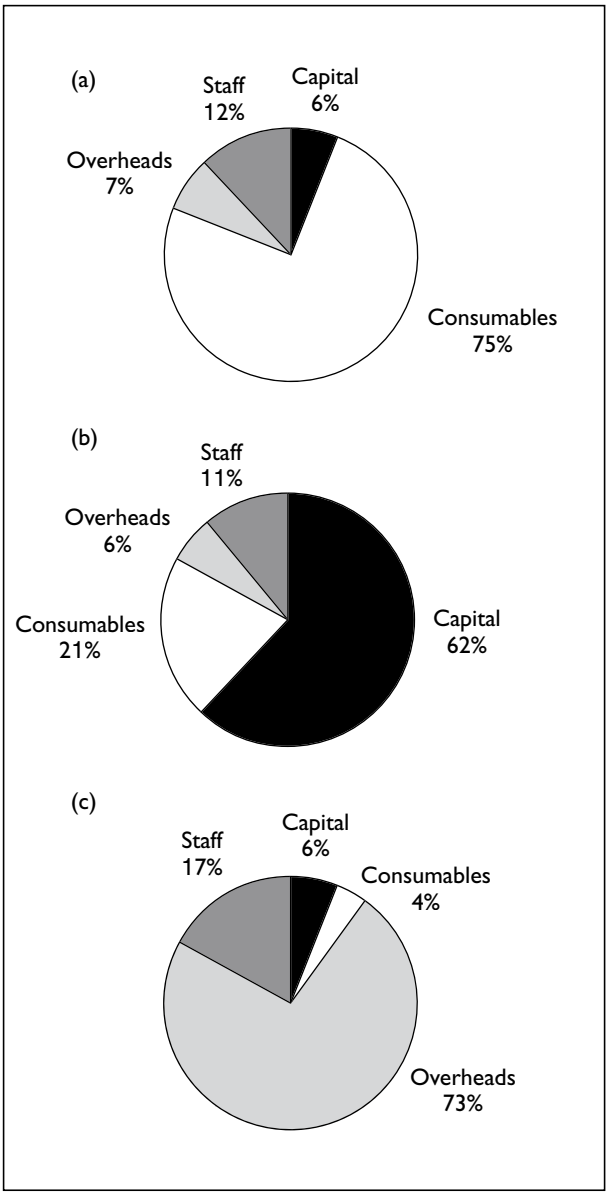


FIGURE 6 *Breakdown of test costs per sample tested for (a) FISH, (b) Q-PCR and (c) karyotyping for throughput of 1000 samples per annum*

TABLE 21 *The influence of throughput on cost per test (£)*

Throughput	UK laboratories with this volume (cumulative %)	FISH		Q-PCR		Karyotyping
		Down syndrome only	5-probe	Down syndrome only	Multiplex	
100	3.3	80.21	113.15	305.42	308.72	120.28
500	25	46.37	79.31	73.19	76.49	93.08
1000[a]	50	42.14	75.07	44.16	47.46	89.68
1500	75	40.73	73.66	34.48	37.78	88.55
5000	100	38.76	71.69	20.93	24.23	86.96

[a] *Approximation to UK median throughput of 1200 cases per annum, adopted for ease of interpretation of results*

[¶] Fixed and variable components, staffing and batch size are held constant at default value.

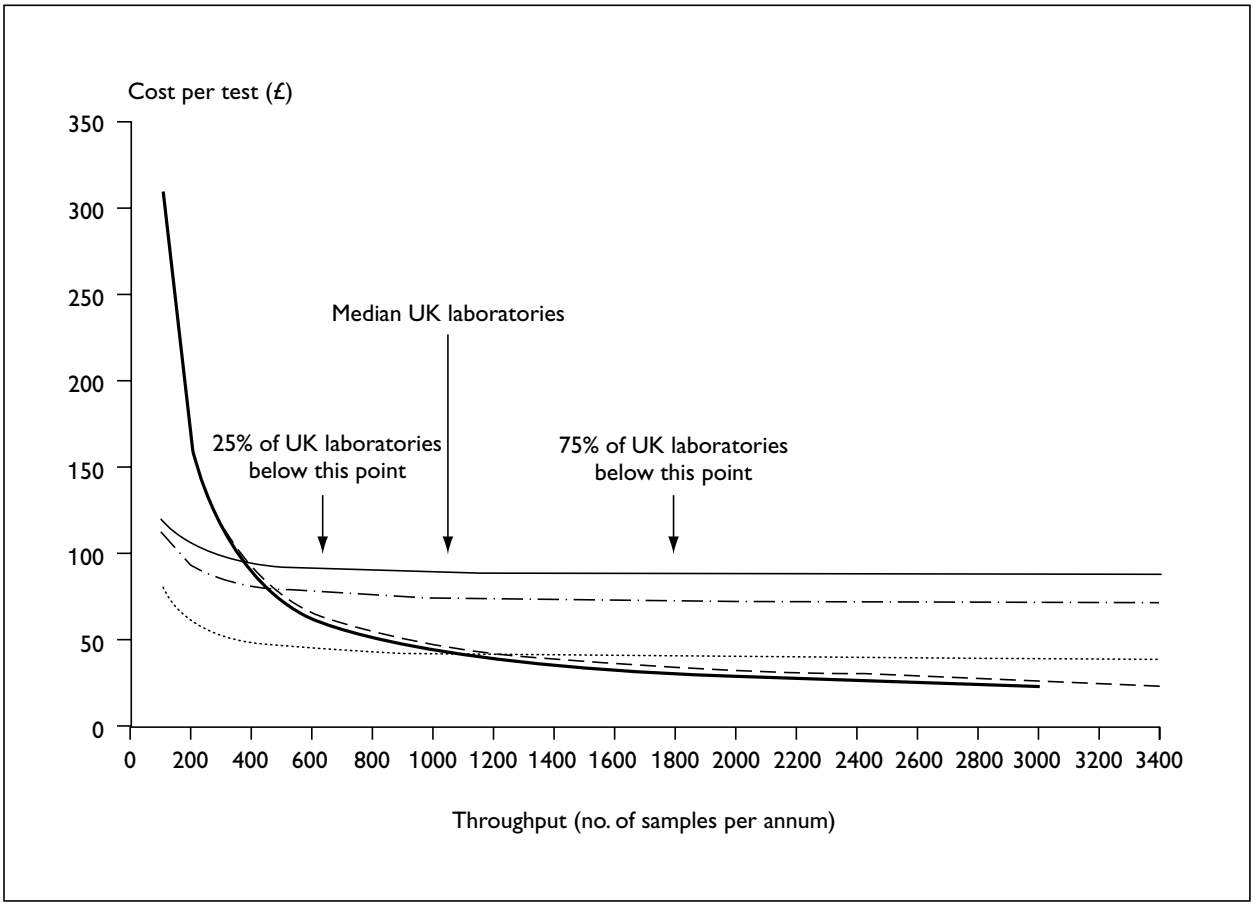


FIGURE 7 *The influence of throughput on cost per test for molecular and traditional prenatal tests (——, karyotyping; — · — ·, FISH 5-probe; ·····, FISH Down only; – – –, Q-PCR multiplex; ——, Q-PCR Down only)*

It is evident from *Figure 7* that for FISH and karyotyping, the cost per test stabilises once a relatively low throughput is achieved (approximately 750), whereas with Q-PCR the cost per test is still falling at a throughput of 3000 tests per annum. The figure also demonstrates that FISH remains less expensive than karyotyping, regardless of the throughput achieved. However, both multiplex and Down only Q-PCR are more costly than karyotyping at laboratory throughputs of less than 400 samples per annum. Similarly, Q-PCR tests only become less expensive than 5-probe FISH once a laboratory processes approximately 450 samples per annum. *Figure 7* shows, however, that more than three out of four laboratories reported throughputs above this level.

Test costs under different laboratory conditions

Ultimately, the cost of molecular tests in the UK service setting will be sensitive to a number of factors other than those discussed above, including:

- changes in equipment and test kit prices
- trends in laboratory organisation (and skill-mix)
- testing approach adopted (i.e. variations in test protocol defining the mix of karyotyping and molecular tests conducted).

Changes in technology, equipment and test kit costs

As demonstrated in *Figure 6,* the cost of a Q-PCR test is highly dependent on the cost of the capital equipment used, whereas the cost of FISH is largely dependent on consumable (test kit) costs.

For Q-PCR, with advances in technology and the possibility of designing and marketing systems for dedicated rather than general purposes, the price of the less sophisticated DNA sequencers required for prenatal testing should fall. Laboratories will therefore be able to trade off reduced price against the sophistication required for research instruments.

In the cost analyses above, the price of a DNA sequencer was assumed to be of the order of £92,000 (based on 1998 prices). If the price of

the sequencer were to fall by 25%, the cost per Q-PCR test would fall by 12% at a throughput of 1000 samples per annum; and if sequencer costs were to halve, the cost of a Q-PCR test would fall by approximately 25%, as illustrated in *Figure 8*.

Similarly, the price of a FISH test kit used in the cost analysis on page 35 ('Labour costs') was the full manufacturer's list price (Vysis AneuVysion kits). For a laboratory service that maximises its purchasing power, either by ordering large quantities or by purchasing in cooperatives, it should be possible to negotiate a discount with the manufacturer. Because consumables (principally test kit costs) account for 75% of the total cost per test for FISH, such purchasing discounts may significantly reduce the cost of testing. For example, for a throughput of 1000 cases per annum, a 10% discount on the cost of the kit would give an approximate reduction of 8% on the cost per test. This relationship is relatively linear, so a 25% cost reduction in test kits would reduce the cost of a test by around 18%.

For the Q-PCR test, fully automated read-out of results became possible with the development of the DNA sequencer. In contrast, FISH has remained non-automated and therefore labour-intensive in terms of pattern recognition. In recent years, considerable effort has been directed at automating the reading of FISH tests. However, although FISH is amenable to a certain amount of automation, very few automated systems are currently on the market, and none has been clinically evaluated.

Trends in laboratory organisation (and skill-mix)

As discussed on page 35 ('Labour costs'), test costs are also sensitive to variations in laboratory organisation and skill-mix. *Figure 9* shows the reported ratio of scientific officers to technical officers versus throughput for the 27 UK laboratories that provided information on staffing in the national survey. As can be seen, for smaller laboratories (< 2000 tests per annum), there is no discernible skill-mix pattern in relation to test throughput, although for the two larger laboratories (4000–5000 tests per annum), the ratio of scientists to technical staff is clearly lower. The reasons for this pattern appear to be both historical and related to the level of specialisation in these laboratories. The two laboratories with the highest throughputs are organisations with high levels of prenatal testing. It can be argued that for higher volume laboratories it is easier to optimise the proportion of technical to scientific staff because senior staff have a full workload at

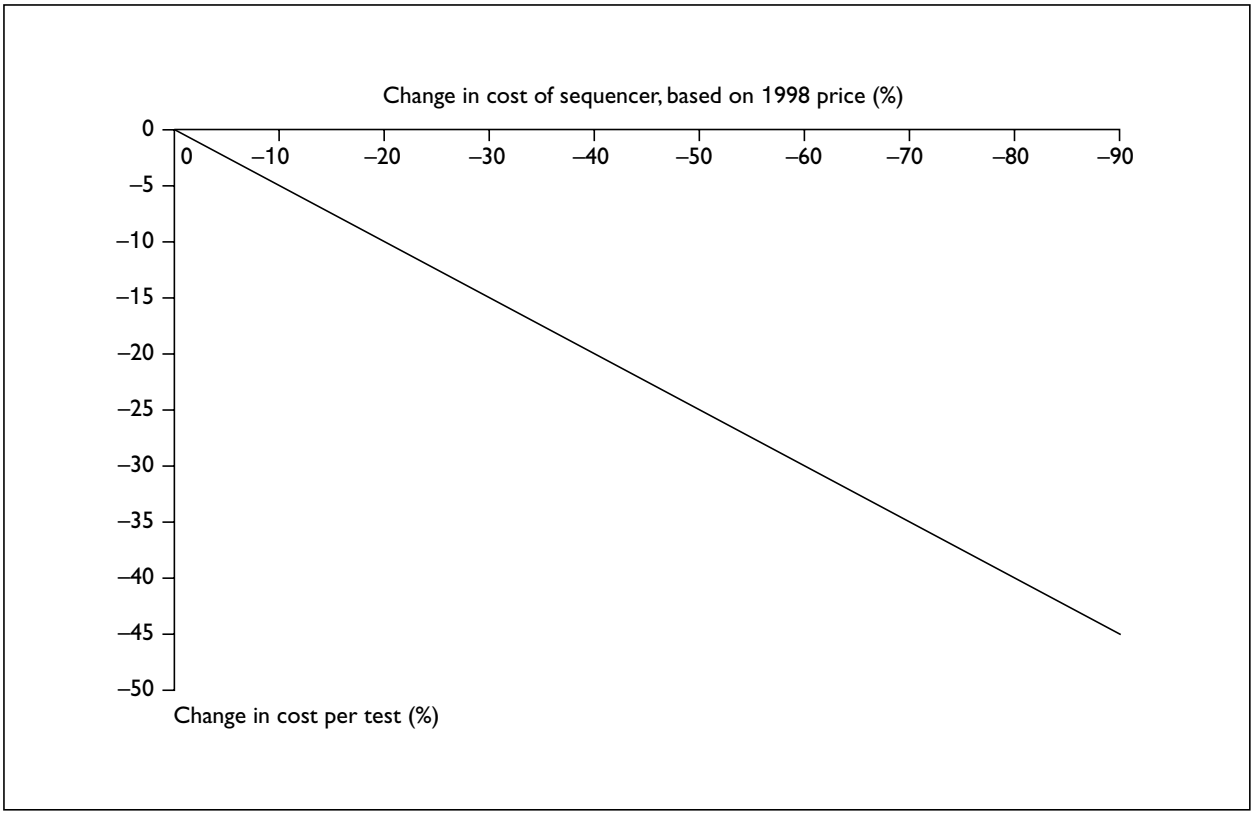


FIGURE 8 *Influence, in percentage terms, on total test cost of change in capital costs for Q-PCR*

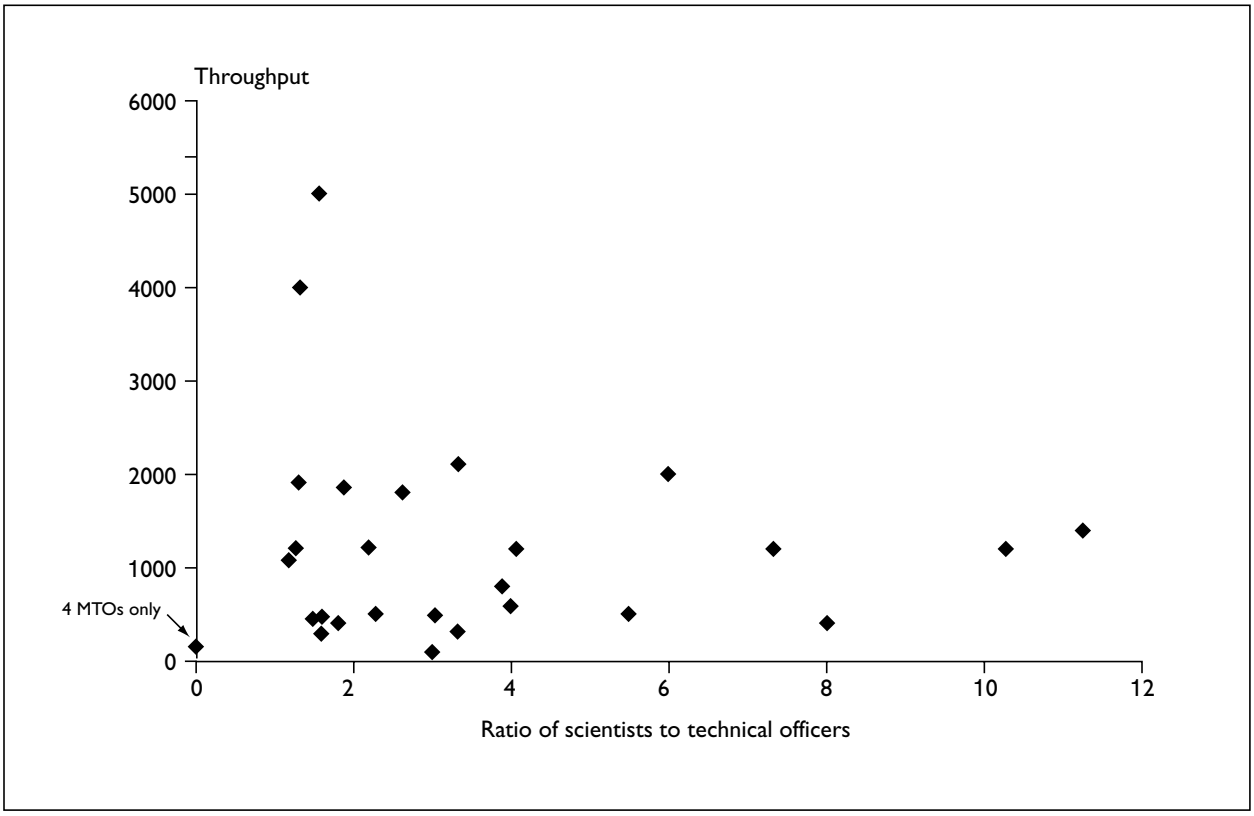


FIGURE 9 *Current skill-mix in different sizes of UK laboratories*

their level of competence without needing to include less specialised tasks that have the potential to be delegated.

In the longer term, any large-scale introduction of molecular tests for prenatal diagnosis should lead to changes in skill-mix within laboratories because many of the tasks associated with molecular tests can be undertaken by more junior staff than are currently employed to assist with karyotyping (*Table 17*).

Figure 10 illustrates the percentage change in cost for the two molecular tests as more senior staff are substituted for MLAs who would be able to undertake most of the simpler preparatory tasks associated with these tests. As can be seen, the cost per test may rise by as much as 33% as higher grade staff are used. This figure also illustrates that FISH is the test most sensitive to changes in staff grades. This might be anticipated because FISH is less capital-intensive and more labour-intensive than Q-PCR (*Figure 6*). In addition, because the Down only FISH test has a proportionately smaller consumable cost element and a higher staff cost than the FISH 5-probe test, it exhibits the highest net percentage increase of the two FISH tests.

Laboratory organisation and skill-mix will be influenced by other routine service commitments, and in such cases the grades and types of staff employed may be more likely to remain laboratory-specific.

Introduction of molecular tests may lead to changes in the organisation of laboratories and rationalisation of laboratory services (particularly if capital-intensive tests are used that also require fewer staff and a different skill-mix).

Discussion

The costing study highlights the fact that the costs of all three types of test are sensitive to local conditions. Furthermore, molecular test costs are also sensitive to the type of test configuration, that is whether the test is designed to only detect Down syndrome (trisomy 21) or whether it detects trisomies 21, 13, 18 and X, Y chromosome abnormalities (5-probe FISH and multiplex Q-PCR). Apart from this, the main factors influencing test cost are laboratory throughput and staff skill-mix. The sensitivity of each test to these various factors will differ depending on the breakdown of a test's costs between capital, labour and consumables. Karyotyping is the most labour-intensive test,

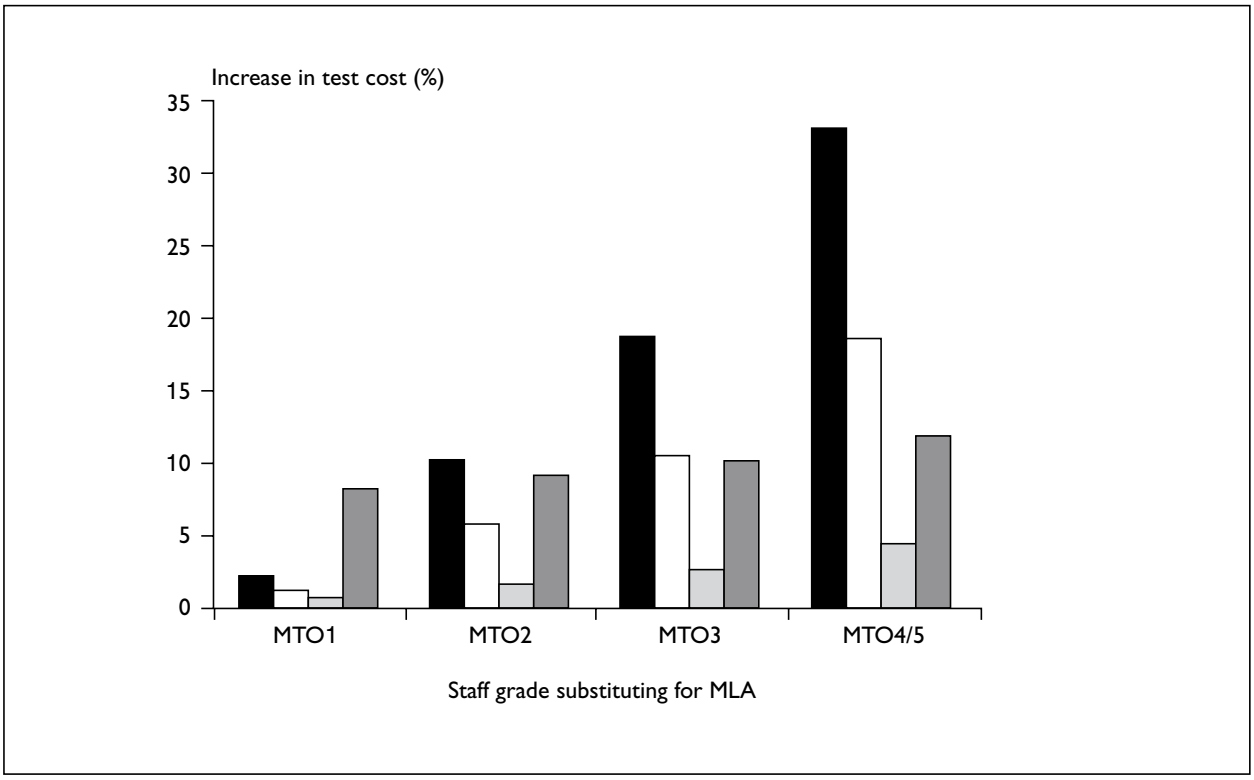


FIGURE 10 *The influence on cost per test of using more skilled staff for routine work (■, FISH, Down syndrome only; □, FISH, 5-probe; ▫, Q-PCR, Down syndrome only; ▪, Q-PCR, multiplex)*

Q-PCR the most capital-intensive, and consumable costs (test kits) are the major part of FISH test costs. The relative costs of these three tests will therefore change under different conditions.

For a laboratory with a throughout of 1000 samples per annum, close to the UK mean reported by laboratories, karyotyping will be the most expensive test to perform, with FISH and Q-PCR (Down only test version) calculated to incur approximately half the direct test costs of karyotyping. This pattern is maintained if a multiplex Q-PCR test were to replace the Down only one. However, use of the 5-probe FISH test would increase test costs substantially, although FISH would still be less expensive than karyotyping.

The cost of all three types of test will decrease with increased throughput to varying degrees (*Figure 7*). However, Q-PCR is most sensitive to throughput, particularly at lower throughputs. This means that both versions of the Q-PCR test will become less expensive than 5-probe FISH in laboratories processing more than 450 samples per annum. Furthermore, multiplex Q-PCR will even become less expensive than Down only FISH by the time a laboratory processes just over 1100 samples per annum. For the 50% of larger UK laboratories processing this number of specimens each year, multiplex Q-PCR would therefore be the less costly option, as well as providing more information than Down only FISH. Multiplex Q-PCR becomes increasingly favourable with increased throughput, as the cost per test continues to fall for Q-PCR even at 3000 samples per annum.

Conversely, for smaller laboratories, FISH would appear to be the preferred option. Both FISH tests (Down only and 5-probe) remain less expensive than karyotyping regardless of the throughput. However, Q-PCR tests (both multiplex and Down only) become more expensive than 5-probe FISH at a laboratory throughput below 450 samples. Approximately one in four laboratories reported throughput of prenatal tests at this level in the national survey. Furthermore, both Q-PCR tests also become more expensive than karyotyping in laboratories with throughputs below 400 samples per annum.

Another factor that might influence these cost patterns is the skill-mix in a laboratory. Karyotyping is most sensitive to skill-mix factors because of the large staff cost element (*Figure 6*). Of the two molecular tests FISH is the most sensitive to changes in staff grade (*Figure 10*). The test costs presented in this chapter are based on a large

laboratory with an appropriate skill-mix for the tasks performed. However, as the UK laboratory survey demonstrated, there does not appear to be a simple relationship between laboratory throughput and reported skill-mix, other than some indication of a lower ratio of scientists to technical officers in larger laboratories processing more than approximately 4000 samples per year (*Figure 9*).

Throughout this exercise we have assumed that equipment used for prenatal testing is not available for other activities within the laboratory. This will have increased estimates of the average cost of capital-intensive tests compared with laboratories where sharing does take place. This assumption may therefore have disadvantaged the Q-PCR test in terms of its relative cost in smaller laboratories.

In summary, therefore, the actual costs of particular testing procedures are likely to be laboratory-specific; they will be dependent on current throughput, staff employed, spare capacity and laboratory organisation.

In future, additional factors will come into play in terms of relative test costs. These include the end of the royalty payment for Q-PCR in 2006, which will reduce test costs by up to £5.40 per test for multiplex Q-PCR. Less easy to predict is the likely impact of automation of FISH. This will reduce labour costs but make the test more capital-intensive, so cost per test may not change substantially.

Finally, no attempt has been made to estimate the cost of changing over from the existing testing situation to any future testing provision that will include molecular tests. It is not known whether laboratories are likely to move wholesale from one test to another and, if not, what combination of testing procedures they are likely to choose. The cost implications of different combinations of tests are explored more fully in chapter 7, once the preferences of obstetricians, women and other stakeholders have been examined in chapter 6.

For the purpose of the evaluation it has also been assumed that the introduction of molecular tests would take place in the same laboratories as those in which karyotyping is currently carried out, with a mean laboratory throughput of just over 1000 samples per annum. However, were some form of laboratory rationalisation to take place this would make molecular tests even more favourable on cost grounds, particularly for capital-intensive tests such as Q-PCR.

So far, the technical capacity and diagnostic accuracy of molecular tests, the capacity for laboratories to undertake these tests, and the direct cost of tests have been explored under different laboratory conditions. However, this does not inform on the higher levels of the Fineberg hierarchy, namely:

- **Diagnostic impact –** does the test replace other diagnostic tests or procedures?
- **Patient outcome –** does the test contribute to improved health or reduced anxiety for the patient?
- **Cost-effectiveness –** does use of the molecular test improve the cost-effectiveness of healthcare compared to alternative interventions?

The information provided by molecular tests is not as comprehensive as that provided by karyotyping, and so one test cannot be a direct substitute for another. In the following chapters we will explore what impact molecular tests might have in terms of their potential to replace or act as an adjunct to the existing karyotyping test, how these tests might be used in combination, and the value of the tests within the healthcare system and to patients and others.

Chapter 5

Methods used to evaluate benefits to parents and other test users

This chapter describes the methods used to ascertain the views of patients and clinicians on the benefits of molecular tests. In addition, the chapter describes a trial of the new molecular tests in a service setting. Details of both of these are summarised at the end of the chapter.

Patient parameters

Tools for assessment of patient parameters

Women who were to receive prenatal testing for chromosome abnormality (following amniocentesis) were recruited to a trial of the new molecular tests from clinics at Trusts in the catchment area of the two genetics laboratories in the study. Participating hospitals were designated either to the control arm of the trial or to the intervention arm. In the intervention arm women received both a molecular and karyotyping test* and in the control arm women received a karyotyping test only. Test preferences, valuation of the test and changes in health status and anxiety were measured for those recruited at amniocentesis and, subsequently, over the period of waiting for the test results. EuroQoL EQ-5D was used to measure health status and the Speilberger State–Trait Anxiety Inventory[64] used to measure anxiety. Measurements were made up to 1 month after receipt of results, because there is evidence that anxiety may persist for this period of time, or longer, especially among younger women who had not previously considered themselves at increased risk.[44]

Questionnaires were piloted on non-pregnant women in the age range 20–40 years, midwives and obstetricians. Questionnaires were then administered to women in both arms of the trial, at specific points during the testing process. *Table 22* shows the sequence of administration and content of each questionnaire. Because

TABLE 22 *Sequence of administration and content of patients' questionnaires*

Questionnaire	Time	Questionnaire content
Women's Questionnaire A (intervention and control)	At booking or clinic visit	Demographic data Preference for test WTP for choice Health status Anxiety (state and trait)
Women's Questionnaire B (intervention only)	Posted with the result of molecular test[a]	Health status Anxiety (state and trait)
Women's Questionnaire C (intervention and control)	Posted with the result of karyotyping[a]	Health status Anxiety (state and trait)
Women's Questionnaire D (intervention and control)	Posted 4 weeks after karyotyping result[a]	Preference for test WTP for choice Health status Anxiety (state and trait)
Partners' Questionnaire	At booking or clinic visit	Demographic data Preference for test WTP for choice Health status Anxiety (state and trait)

[a] *Normal result only*

* Where sufficient amniotic fluid was available to undertake both. The trial did not require extra fluid to be withdrawn.

ethics approval required that patient identity be concealed from the non-scientist members of the research team, all questionnaires were coded and administered by staff in the cytogenetics laboratories. As it was normal practice to mail results of normal tests to patients, Questionnaires B and C were posted with test results, others were posted separately. Patients were therefore responding to Questionnaires B and C with knowledge of their test result.

The first three patients recruited to the study were all accompanied by their partners and each partner asked whether a similar questionnaire could be provided for them to complete. Subsequently, all partners were offered a questionnaire at the booking/clinic visit, based on Questionnaire A for the women.

Table 23 describes the content of the questionnaires listed in *Table 22* in greater detail, including the instruments used. The demographic questions, information leaflet accompanying the questionnaire, and questions other than EuroQoL EQ-5D and Speilberger questions (for which standard formats were used) are reproduced in appendix 3.

TABLE 23 *Patient data and instruments used*

Questionnaire content	Brief details
Demographic data	Age, number of previous pregnancies, ethnic origin, occupation
Preference for test	Constrained choice between molecular test or karyotyping
WTP for choice	Strength of preference and valuation of test
Health status	EuroQoL EQ-5D
Anxiety	Speilberger State–Trait Anxiety Inventory

Recruitment of patients

Patients for the trial were recruited from women who attended amniocentesis clinics during the 3-month period from April to July 1999. Exclusion criteria for the study included previous miscarriages, previous history of a baby with chromosome abnormality, women with serious depressive illness, and those requiring emergency amniocentesis outside the clinic structure.

In Northern Ireland, patients were recruited centrally because the majority of women pass through a single clinic. All women from Northern Ireland were entered into the intervention arm of the study as it was not considered ethical to recruit women to both arms of the trial within one Trust. Women were recruited consecutively and all were offered 5-probe FISH as the molecular test.

In the West Midlands, patients could not be recruited centrally because of the manner in which the service is organised (see page 30, 'Study sites: structuring of prenatal testing services'). Thus, patients were recruited consecutively from three centres (South Warwickshire General Hospitals NHS Trust, Walsall Hospitals NHS Trust and Royal Shrewsbury Hospitals NHS Trust), with the agreement of all consultants at these hospitals. These patients were all offered multiplex Q-PCR as the molecular test during the 3-month recruitment period. In one centre there was a 2-week period when the counselling midwife was on annual leave and only three patients were seen and given amniocentesis during this time. In one centre, patients were recruited for only 6 weeks because of delays in obtaining local ethics committee approval.† Control patients were recruited consecutively from these West Midlands centres immediately after cessation of recruitment to the intervention arm.

All women attending for amniocentesis were invited to take part in the study at the booking visit, during which amniocentesis was arranged. For those patients who were referred from Community Midwives, recruitment was at the counselling visit. The objectives of the study were explained to the patient (and their partner, if attending together) and written consent was obtained. An information leaflet was provided for patients and clinic interpreters were available if required. For women recruited to the intervention group, it was explained that in order to perform both tests an adequate volume of amniotic fluid was required, and women were informed that the molecular test would only be performed if sufficient amniotic fluid remained after that required for karyotyping had been removed. Patients were not given an explicit time for receipt of their molecular test result. Women were also assured that if an abnormality were to be detected by the molecular test their obstetrician would be

† Unresolved procedural issues relating to Multicentre Research Ethics Committee and Local Research Ethics Committee authority.

informed immediately.[‡] In these cases the patient was withdrawn from the study.

Measuring parental anxiety levels

In terms of parental anxiety, it can be speculated that an important trade-off for women and their partners will be the fact that although a molecular test can produce a quicker result it will provide no information about rarer abnormalities detectable by karyotyping. Thus, although there may be benefit arising from the reduction of anxiety in the short term, this may be offset by increased anxiety arising from uncertainty about the completeness of the information given by the test result.[65] To quantify women's experiences of molecular tests compared with the traditional karyotyping test, anxiety levels were measured throughout the testing process for both groups of women; those offered a molecular test and those following the traditional testing process (controls). Anxiety was measured with the short form of the Speilberger State–Trait Anxiety Inventory. This instrument allows differentiation of the specific anxiety caused by current circumstances (State) from the normal anxiety levels of the person tested (Trait).

For the intervention group, the sequence of administration of the questionnaires is shown in the top part of *Figure 11*. Four questionnaires were administered: at booking/amniocentesis (Q_{I1}); at molecular test result (Q_{I2}); at karyotyping result (Q_{I3}); and 4 weeks after karyotyping (Q_{I4}). For the control group, as shown in the bottom part of *Figure 11,* as well as the questionnaire at booking/amniocentesis, questionnaires were completed at karyotyping result (Q_{C2}) and 4 weeks after this (Q_{C3}).

Patient population entered in study

In total, 67 patients were recruited to the intervention group in Northern Ireland and 74 in the West Midlands. Only three women refused to take part in the study (Northern Ireland). In the West Midlands, 53 patients were recruited to the control group; none refused. Eighty-nine partners accompanied patients for amniocentesis or the booking visit for amniocentesis. Four partners refused to fill in questionnaires and 28 (31.5%) refused to supply demographic data on age and income band.

Table 24 shows recruitment to the intervention and control groups, the age range of women and of partners, and the mean number of pregnancies for each woman.

The majority of the patients were white/Caucasian (92%) with 5% women of Asian origin, 1% Afro-Caribbean and 2% non-responders. For 24% of the women recruited this was their first pregnancy, 22% their second, and 23% their third; of the remainder, 23% had three or more pregnancies and 8% were non-responders. There were no significant differences in age or ethnic origin between women in the intervention group and control group (Mann–Whitney, $p = 0.97$ and $p = 0.06$, respectively), but women in the intervention group had a slightly lower mean parity at recruitment (mean = 0.85) than the control group (mean = 1.2, $p = 0.02$).

Questionnaire responses

Table 25 shows the pattern of completion for the four questionnaires. There were no significant differences between the intervention and control groups of women in terms of completion of questionnaires. All results were calculated using the number of respondents for each questionnaire.

Clinician surveys

General surveys of obstetricians

In order to explore obstetricians' views about the test choices they would be likely

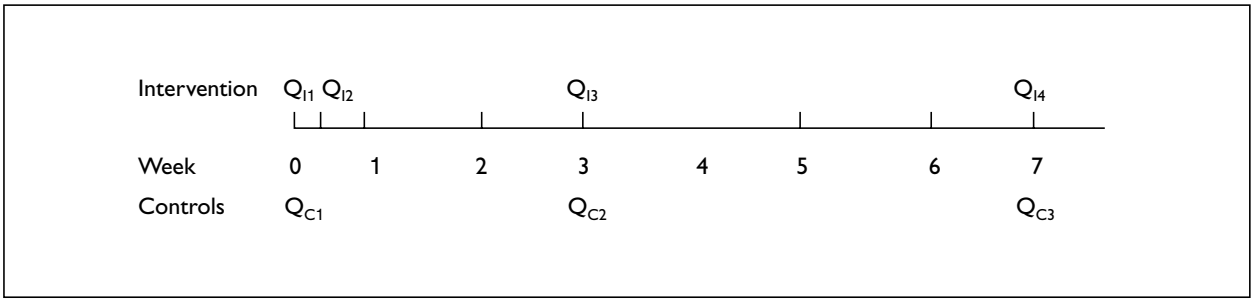


FIGURE 11 *Survey administration points throughout the testing process*

[‡] Where a positive molecular test occurred, karyotyping was given high priority by the laboratory and the sample was 'fast-tracked' at the earliest opportunity, once sufficient cells were in division.

TABLE 24 *Patient recruitment to the intervention and control groups and demographic data*

Trial arm	Location	Women				Partners		
		No.	Age range	%	Mean previous pregnancies	No.	Age range	%[a]
Intervention	West Midlands	74	16–19	4.0	1.34	50	16–19	4.5
			20–24	4.0			20–24	6.8
			25–29	16.2			25–29	15.9
			30–34	25.7			30–34	22.7
			35–39	41.9			35–39	40.9
			40–44	6.8			40–44	6.8
			45 and over	1.4			45 and over	2.4
	Northern Ireland	67	30–34	4.5	1.89	38	30–34	0
			35–39	64.2			35–39	53.0
			40–44	29.9			40–44	23.5
			45 and over	1.5			45 and over	23.5
Control	West Midlands	55	25–29	3.8	1.65	46	25–29	10.9
			30–34	26.4			30–34	23.9
			35–39	52.8			35–39	45.7
			40–44	17.0			40–44	15.2
			45 and over	0			45 and over	4.3

[a] *30.7% non-responses to demographic questions from partners*

TABLE 25 *Numbers of women completing consecutive questionnaires*

Questionnaire	Women		Total
	Northern Ireland	West Midlands	
A (recruitment)	67	74	141
B (molecular result)	56	57	113
C (karyotyping result)	47	49	96
D (4 weeks post C)	44	47	91

to make were molecular tests to become more generally available, all obstetricians in the West Midlands and Northern Ireland were surveyed. This survey had three objectives:

- to estimate the degree to which formal protocols were used for selecting patients
- to ascertain views on the likely demands on services arising from introduction of molecular tests
- to explore the possible choices obstetricians would make for chromosome tests on amniocytes and chorionic villous samples.

Following presentation of interim results from this survey, the HTA Programme Visiting Panel asked for the questionnaire to be disseminated nationally to obstetricians in all UK Trusts. From both surveys, responses were received from 454 obstetricians (35.4%). This relatively low response rate was mitigated by indications from respondents that they were responding on behalf of their colleagues as lead clinician for prenatal testing and amniocentesis in their Trust. At Trust level, questionnaires were sent to 314 Trusts and responses were received from 250 (79.6%), including the West Midlands and Northern Ireland. Response rates from individuals within the West Midlands and Northern Ireland were higher (perhaps because involvement in the present HTA project had provoked discussion of the issues at Regional Obstetric meetings) with 70.2% of individuals responding in Northern Ireland and 59.4% responding in the West Midlands. Responses for the two surveys were collated and are presented in the following chapters for all UK obstetricians unless otherwise stated.

Survey of obstetricians related to individual patients

In addition to the above surveys, during the first stage of the study both laboratories (Northern

Ireland and West Midlands) included a short questionnaire with all abnormal karyotyping test results sent to obstetricians. This questionnaire asked respondents to identify what, if any, clinical benefits would have resulted from a more rapid test result for this sample of their own patients. A similar-sized, random sample of normal results was followed up in the same manner. In this way information was collected from all obstetricians receiving karyotyping results about the choices they would have made for these individual patients had a molecular test been available at the time of amniocentesis.

In total, 200 questionnaires were despatched, 100 from each laboratory; 109 responses were received – 62 from obstetricians in Northern Ireland and 47 from obstetricians in the West Midlands (overall 55% response rate). Ninety responses were for cases with a normal result and 19 were for cases where an abnormality had been detected. There were no significant differences in age (range 24–44 years) for women in the normal and abnormal groups; details of the abnormal karyotyping results were not available to the research team.

Midwives

From the national laboratory survey it is evident that midwives are an important source of information for patients in many Trusts and so a survey of midwives was also carried out in the study regions. In Northern Ireland, midwives were not involved in counselling, instead clinical geneticists performed this service. Therefore, the survey was limited to midwives in the West Midlands study region with specific training or an interest in prenatal chromosome analysis. These midwives were surveyed using a questionnaire similar to that used in the obstetricians' survey. The questionnaire was mailed to 21 specialist midwives and 21 responses were received (100% response rate).

Health commissioners surveys

Health commissioners are also stakeholders in the debate about prenatal testing. In particular, they have a legitimate perspective in terms of potential alternative uses for any resources invested in prenatal molecular testing. A questionnaire was therefore developed to explore their preferences.

This questionnaire offered a constrained choice between molecular tests and karyotyping, and it was sent during May 1999 to all 105 Directors of Public Health in England and Wales. It was first piloted by two Directors of Public Health and two Consultants in Public Health Medicine.

Seventy-six (72.4%) of those mailed either responded personally or passed the questionnaire to the relevant colleague for completion.

Summary

In order to ascertain and, in some cases, quantify users' preferences for molecular tests, three stakeholder groups were identified – women tested, clinicians ordering the tests and those who commission services, represented by Directors of Public Health or Consultants in Public Health Medicine. Subsequently partners accompanying women to amniocentesis clinics also requested that their views be surveyed.

Women and their partners were recruited to a prospective study in the two regions served by the participating genetics laboratories. Recruits to the intervention arm (141 women; 74 in the West Midlands and 67 in Northern Ireland) received molecular and traditional tests and answered four sequential questionnaires, as shown in *Figure 11*. The first (at amniocentesis or booking, whichever was the earliest) and last questionnaires examined test preferences and valuation of the tests. All four questionnaires recorded anxiety (Speilberger State–Trait Anxiety Inventory) and health status (EuroQoL EQ-5D). In the control arm, women (55 cases, West Midlands) received a traditional test only and answered three questionnaires (i.e. omitting the molecular test result questionnaire). All control questionnaires contained health status and anxiety questions. Demographic data were collected from both arms of the study.

Health commissioners and two groups of clinical staff were surveyed; obstetricians were surveyed nationally (including Scotland, Wales and Northern Ireland) and midwives were surveyed in the West Midlands only. The response rate from obstetricians was 35.4%, representing 80% of Trusts in the UK. The response rate from West Midlands midwives was 100% and from health commissioners 72.4%. The results of all these surveys are presented in the following chapters.

Chapter 6

Patient and physician choices

Introduction

The assessment of technical capacity and diagnostic accuracy of molecular tests (Fineberg hierarchy levels 1 and 2) has been reported in earlier chapters, particularly chapter 2. This assessment involved a comparison of the performance of the two types of molecular test against the accepted gold standard (karyotyping). As discussed in chapter 1, the likely impact of molecular tests – the extent to which molecular tests will replace other diagnostic procedures – must also be considered (Fineberg level 3) before any higher level evaluation can be conducted.

As reported in chapter 2, molecular tests can be configured either to detect Down syndrome (trisomy 21) only or to detect trisomies 21, 18, 13 and X, Y chromosome abnormalities. Thus, information is required from clinicians on molecular test configurations most likely to be requested. It is evident that the manner in which molecular tests are used will be dependent on several factors, including current practice for patient selection for amniocentesis and clinicians' preferences and perceptions about particular tests. Also relevant will be how possible options for prenatal tests are managed within individual Trusts and laboratories, and how choices are discussed with patients.

Therefore, as a precursor to the higher levels of evaluation, this chapter explores the preferences of clinicians and other stakeholders for test configuration, testing regimes and pattern of use of both molecular tests and karyotyping. As well as future test preferences, surveys of obstetricians and midwives also recorded current practice in patient selection and counselling. In addition, other surveys were used to explore women's preferences and those of other stakeholders. The results of these surveys are reported in this chapter. Once the likely impact of a molecular test is assessed, patient outcome and cost-effectiveness will then be considered (Fineberg hierarchy levels 4 and 5).

Results of regional and national survey of obstetricians

Current patient selection for amniocentesis

UK obstetricians were asked about the current selection process for amniocentesis. *Figure 12* illustrates the range of age cut-off points reported for that part of risk assessment related to maternal age. Responses indicated significant differences between Trusts in the age cut-off points used. Furthermore, only 103 Trusts (26.4%) reported the existence of a formal written protocol; 28.4% had a jointly agreed informal protocol; and in 45.2% of Trusts individual clinicians made their own choice. Where protocols were in place, only one-third (36%) of respondents reported that adherence to these was rigidly maintained.

Absence of a Trust-wide protocol for selection of patients makes it difficult to predict how tests will be used locally and, once introduced, it will be hard to monitor appropriate and effective use of such tests.

Pre-test counselling and giving results

In most responding Trusts, it was reported that either obstetricians or midwives performed counselling before amniocentesis. For 91% of Trusts, obstetricians were the main counsellors. In only 20% of responding Trusts were genetic counsellors specifically employed, usually clinical scientists or clinicians with a special interest in clinical genetics. In 53% of Trusts, midwives were involved in counselling, although only 39% of these midwives had genetic counselling qualifications; in 5% of Trusts specialist midwives were responsible for **all** counselling. A majority of obstetricians, 57%, considered that the introduction of molecular tests would not place extra demands on counselling, although 27% remained unsure about this.

National survey respondents were also asked about current practice for giving karyotyping results to patients. Reported practice for giving normal (i.e. negative) results is shown in *Table 26*. Only 25% of respondents considered that the way in which results are currently given to women

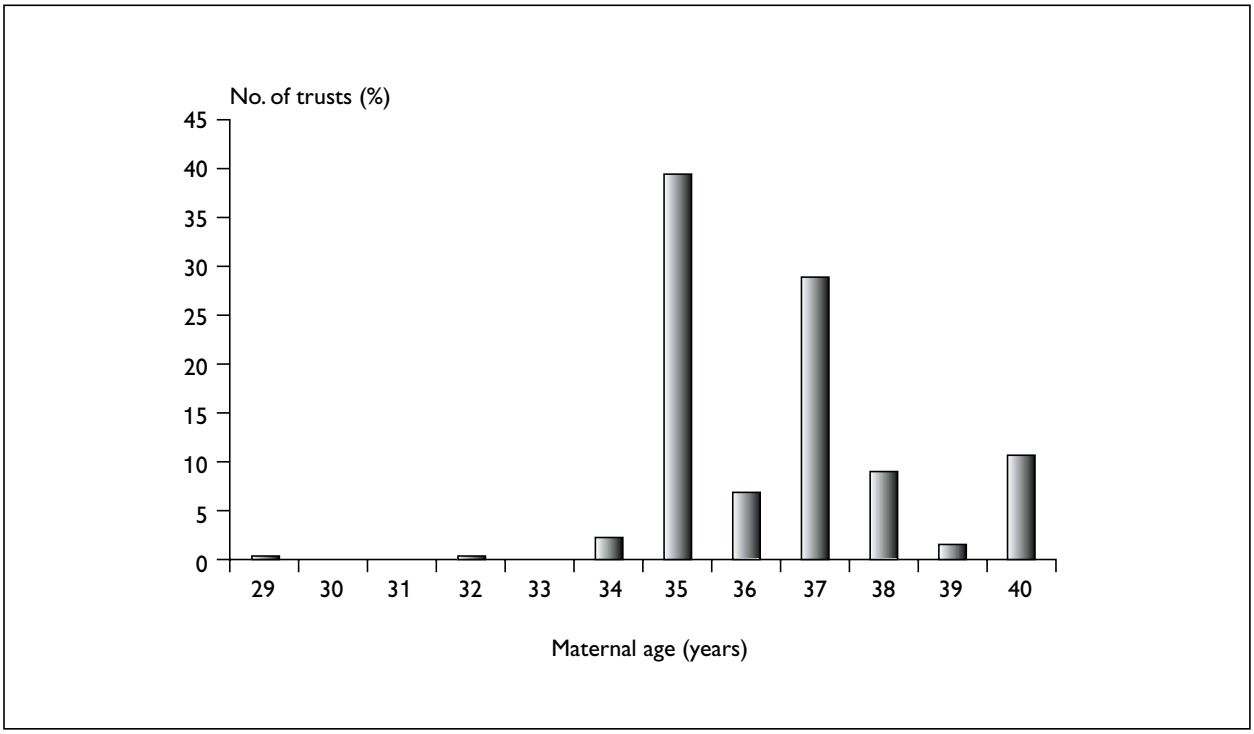


FIGURE 12 *Range of age cut-off points for selection for chromosome analysis*

TABLE 26 *Current practice for giving patients result where test has not detected abnormality*

Method of giving normal results[a]	UK Trusts (%)[b]
Appointment made	24.2
Patient asked to phone antenatal clinic	59.6
Patient asked to phone laboratory	0.6
Letter from antenatal clinic	22.7[c]
Letter from genetics laboratory	9.6
Other	4.3

[a] *A few services tailor giving results to patient preference*
[b] *More than one method employed in most Trusts*
[c] *91% using 1st class stamp, 9% using 2nd class stamp*

would change with the introduction of molecular tests.

Obstetricians were also asked what they told their patients about how long the test would take. Responses are shown in *Figure 13*. The majority told women to expect a result within 15–21 days.

Finally, for the national survey only (excluding West Midlands and Northern Ireland, where the answer was already known, see page 30, 'Study sites: structuring of prenatal testing services'), a question was asked about the current practice for communicating abnormal results to patients. In almost all cases, the laboratory phoned either the midwife or the obstetrician. Most obstetricians (75.9%) and most midwives (88.5%) then phoned their patients; 18% of obstetricians followed this up with a letter. A small number (1%) of obstetricians relied on the laboratory to give patients abnormal results.

Obstetricians' test preferences

Obstetricians were also asked a series of questions about their ideal preferences in terms of which chromosome abnormalities a molecular test should ideally detect, assuming that molecular tests could be extended to test selectively for other conditions in addition to Down syndrome. In the first instance obstetricians were given a constrained choice and asked to assume that a molecular test would replace karyotyping. They were then asked to choose between three options: (i) a test for Down syndrome only (trisomy 21); (ii) a test for Down syndrome, trisomies 13, 18 and numerical sex chromosome abnormalities; (iii) a test for Down syndrome plus the option of selecting any from (ii).

In total, 58 obstetricians (58/454, 12.8%) failed to respond to any part of the questions on test preferences, so the following percentages are calculated in terms of the 396 respondents. A majority of

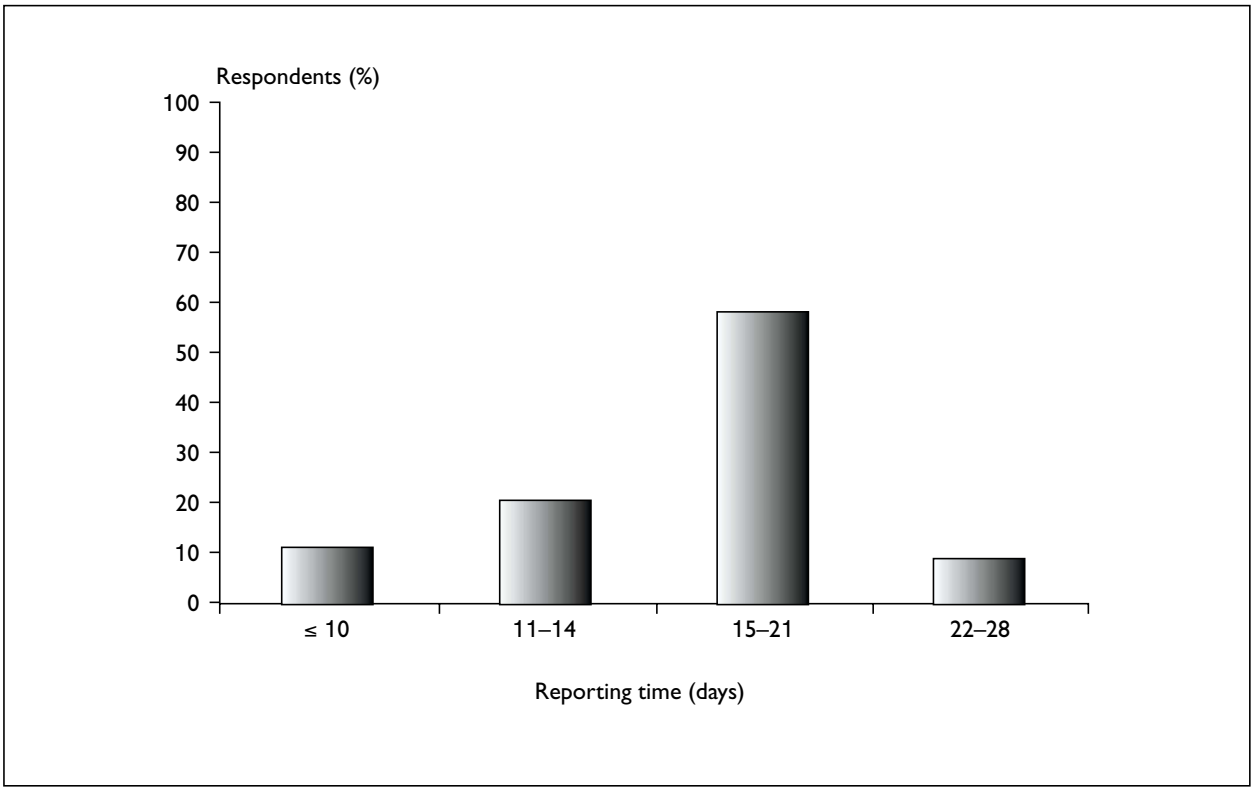


FIGURE 13 *Obstetricians' advice to patients on length of wait for karyotyping result*

obstetricians (57.3%) expressed a preference for option (ii), with only 7 respondents (1.8%) favouring testing for Down syndrome only. The remaining 41% stated a preference for the third option.

In a further question, obstetricians were asked to assume that, in addition to trisomy 21, a molecular test could be ordered for any of the abnormalities specified in *Table 27*. In cases where an individual patient was **not** identified to have a specific risk prior to testing (arising from past medical history, family history or abnormal ultrasound result) obstetricians were asked the following questions about their preferred molecular test configuration:

- which abnormalities they considered it 'essential' that the molecular test should be able to detect
- which abnormalities they considered only 'desirable'
- which abnormalities they considered were 'not required'.

Approximately half the obstetricians reported that it would be essential for the molecular test to be able to detect trisomies 18 and 13, in addition to trisomy 21. A small minority (5–7%) considered trisomy 18 and 13 tests were not required. However, the pattern of responses differed significantly ($\chi^2 = 26.9$, $p < 0.001$) with regard to tests for numerical sex chromosome

TABLE 27 *Obstetricians' choices for structure of molecular tests*

Abnormality	Essential (%)	Desirable (%)	Not required (%)	Don't know (%)	Missing responses (%)
Trisomy 18	209 (52.8)	143 (36.1)	23 (5.8)	11 (2.8)	15 (3.8)
Trisomy 13	197 (49.7)	144 (36.4)	30 (7.6)	15 (3.8)	15 (3.8)
XXY (Klinefelter) syndrome	28 (7.1)	139 (35.1)	144 (36.4)	74 (18.7)	15 (3.8)
XYY syndrome	24 (6.1)	118 (29.8)	139 (35.1)	105 (26.5)	15 (3.8)
XXX syndrome	24 (6.1)	128 (32.3)	136 (34.3)	96 (24.2)	17 (4.3)
45X (Turner's) syndrome	59 (14.9)	147 (37.1)	120 (30.3)	60 (15.2)	15 (3.8)

abnormalities. A small minority (6–7%) considered detection of numerical sex chromosome abnormalities was essential, 15% replying 'essential' for Turner's syndrome. For the remaining 85% of respondents, opinion was evenly split between 'desirable' and 'not required'. Moreover, significantly more clinicians responded 'don't know' ($\chi^2 = 35.4$, $p < 0.001$) to the question on numerical sex chromosome abnormalities. These abnormalities generally have more variable and milder effects than the trisomies (*Table 1*).

Obstetricians were also asked to reflect on their likely pattern of test requests (unconstrained) if molecular tests were to be introduced into the service setting. The responses of 309 (82.8%)* respondents who replied to this question are shown in *Table 28*. The leading preference, for over half the obstetricians (56.8%), was for the majority of patients to receive a molecular test, with karyotyping reserved for a minority. Over half these respondents (58%) also reported that they would request a molecular test able to detect trisomies 21, 18 and 13 and the numerical sex chromosome abnormalities.

Fewer than 30% of obstetricians considered they would be likely to request **both** tests for some of their patients, but only 15% would request both tests for all, or the majority of, women. Two-thirds of those who would request both tests for their patients reported that they would choose the molecular test that detected the widest range of abnormalities, although maximum diagnostic information would, in the long run, be available to these patients as they were also receiving karyotyping.

Obstetricians were further asked to reflect on their likely pattern of test requests if they were constrained to choose only **one** test for their patients and did not have the option of requesting both the molecular test and karyotyping. Test preferences for constrained choice are shown in *Table 29*. Overall, less than one in ten (8.8%) would restrict all their testing to karyotyping. With constrained choice, 39.6% would choose molecular tests for the majority of their patients and 23.9% would allow the patient to choose. However, approximately one in four (27.3%) were unsure of their preference.

A minority (15.8%) of obstetricians reported that they already had access to molecular methods for some cases, most as an add-on test to karyotyping. Not all clinicians in the same Trust, however, requested this service. Of the clinicians able to request molecular tests, 98% had access to FISH, and 50% of these also reported that they requested FISH on chorionic villus samples.

Finally, obstetricians were also asked about the necessity to include a test configuration capable of detecting some of the less common abnormalities (e.g. unbalanced and balanced† rearrangements, see *Table 1*). Hypothetical options included routine

TABLE 28 *Obstetricians' views on their pattern of test requests (unconstrained)*

Preference for type of test	% expressing this preference			
	Total	Down only	Trisomies 21, 18 and 13	Trisomies 21, 18, 13 and X,Y chromosomes
Karyotyping for the majority, molecular tests for the minority	14.5	2.0	4.3	8.2
Molecular tests for the majority, karyotyping for the minority	56.8	5.6	19.0	32.2
Both tests for all women	8.5	0.6	2.0	5.9
Both tests for the majority of women, karyotyping for remainder	6.5	1.0	1.6	3.9
Both tests for the minority of women, karyotyping for remainder	13.7	0.3	6.6	6.8

* This question was inserted into the national survey as the questions used in the prior survey of West Midlands and Northern Ireland obstetricians were refined. Thus, data is analysed with a reduced denominator.

† Balanced rearrangements have no effect on the fetus other than an increased risk of the child, once an adult, having chromosomally unbalanced offspring.

TABLE 29 *Test preferences where obstetrician constrained to one test only*

Preference for type of test constrained to one test	% expressing this preference			
	Total	Down only	Trisomies 21, 18 and 13	Trisomies 21, 18, 13 and X,Y chromosomes
Molecular tests for majority of women, karyotyping for a minority	39.6	2.4	13.1	24.1
All patients freedom to choose which test	23.9	1.9	2.7	19.3
Karyotyping only for all women	8.8			
Did not answer question	27.3			

testing for these less common abnormalities in all higher-risk cases (see page 18, 'Relative sensitivity and specificity of molecular tests'), regardless of the abnormality for which they might be judged to be at risk; or only testing in those perceived to be at high risk of this type of abnormality. *Table 30* shows their responses.

The vast majority of obstetricians (66.3–70.2%) were in favour of testing only specific high-risk cases for these chromosome abnormalities; only 16.4–21.6% would wish to routinely test all higher-risk cases; and approximately one in ten (10.5–13.4%) were unsure. In cases where a high risk of a particular fetal chromosome abnormality had been identified, two-thirds of obstetricians (263/396, 66.4%) indicated that they would prefer to ask for a molecular test in addition to karyotyping; just over a quarter (26.5%) would request only karyotyping; 6.0% reported they would not ask for karyotyping.

Finally, survey respondents were asked whether they were currently able to obtain prenatal chromosome tests for all those women judged to need genetic testing, and are suitable for amniocentesis. Over 97% of obstetricians indicated that supply was able to meet demand and a similar response was received from West Midlands midwives.

Obstetricians' test choice for individual patients

In addition to the national survey, obstetricians in the two study areas were also asked retrospectively about a number of their own patients (see page 48, 'Survey of obstetricians related to individual patients'). They were first asked to identify which selection criteria they had used to select each patient for prenatal testing. *Table 31* shows the replies for 109 cases broken down by reason for amniocentesis and type of karyotyping result (normal or abnormal). Free text reasons for undertaking amniocentesis for this group of patients are given in appendix 4. Where appropriate these have been incorporated in *Table 31*. This demonstrates that, in this small group of patients, only those with increased thickness of the translucent area of nuchal fold were likely to have an abnormal karyotyping result ($\chi^2 = 6.61$, $p = 0.01$). Other selection criteria showed no significant correlation with karyotyping results.

The opportunity was taken, during this smaller survey, to explore obstetricians' opinions on their likely use of molecular tests for this specific patient group, rather than the hypothetical body of patients considered in the previous section. Obstetricians indicated, retrospectively, that they

TABLE 30 *Obstetricians' preferences for information about structural chromosome abnormalities*

Obstetricians request	Extra marker[a] (%)	Unbalanced[b] (%)	Balanced[c] (%)
Routinely in all higher-risk cases	84 (21.6)	75 (19.3)	64 (16.4)
Only for those perceived to be at high risk of this type of abnormality	258 (66.3)	273 (70.2)	275 (70.3)
Not sure	47 (12.1)	41 (10.5)	52 (13.3)

[a] *Information supplied in survey: "Extra marker chromosome, prevalence less than 1 in 2500 cases tested"*
[b] *Information supplied: "Unbalanced structural rearrangement, prevalence less than 1 in 10,000 cases tested"*
[c] *Information supplied: "Balanced structural rearrangement, prevalence less than 1 in 10,000 cases tested"*

TABLE 31 *Selection criteria for amniocentesis for 109 retrospective cases in the obstetrician survey in two study areas*

Reason for amniocentesis	% of group to which selection criteria apply[a]			Total selections
	With abnormal karyotyping result	With normal karyotyping result	Difference between normal and abnormal	
Positive serum test	52.6%	43.8%	NS	45
Previous child with genetic abnormality	5.3%	10.0%	NS	9
Family history	0%	6.3%	NS	5
Elderly gravida	26.3%	55.0%	NS	48
Significant nuchal translucency	26.3%	6.3%	$\chi^2 = 6.61, p = 0.01$	8
Abnormal ultrasound scan	10.5%	5.0%	NS	16
Maternal anxiety	5.3%	2.5%	NS	4
Total cases	19	90		109

[a] *Some cases are in more than one category*
NS, not significant

would have requested a molecular test in eight out of ten (79%) cases. There were no significant differences in the test choices that clinicians reported for patients in the two groups (i.e. comparing those with normal and abnormal karyotyping results). For each patient, obstetricians were also asked about the importance of (i) a test to detect the five most common chromosome abnormalities that could, at that time, be detected using molecular tests[‡] and (ii) a test that could also detect all other abnormalities that could not be detected currently by molecular tests.[§]

First, obstetricians were asked, if molecular tests were to replace karyotyping, what molecular test option they would have chosen for each patient. Responses indicated that one in five (20.8%) would have chosen a Down syndrome only molecular test, and nearly two-thirds (62.3%) would have ordered a molecular test for the five most common abnormalities. A significant minority (17%), however, would have preferred a test request format that enabled the components of the test (Down plus other abnormalities) to be specified for each patient. Of this last group, 90% would have ordered trisomy 18 as an addition to trisomy 21, and 80% would have ordered trisomy 13 as an addition to trisomy 21; both chromosome abnormalities lead to severe disability and high mortality in the first few weeks. Only 15% or less would have also ordered tests for numerical sex chromosome abnormalities where the impact on the child in these cases is generally milder and more variable (*Table 1*). These responses broadly replicated those from the wider general survey, not linked to test requests for individual patients, reported in the previous section.

Secondly, obstetricians were given data about incidence of rarer abnormalities[¶] and were once again offered the following (unconstrained) choice for each particular patient:

- karyotyping only
- molecular test only
- both karyotyping and molecular test.

Table 32 indicates that with this information on rarer abnormalities, obstetricians still reported that they would have selected the karyotyping only option in only one in ten cases (10.1–11.9%). In the remaining cases, the obstetrician would have chosen a molecular test, but responses were equally divided in terms of whether a karyotyping test would also have been requested. It should be noted that molecular tests are unlikely to be able to detect these abnormalities in the near future.

Finally, in terms of any need for additional counselling associated with molecular tests, 60% of obstetricians from this survey considered that for these particular patients molecular

[‡] Trisomies 21, 18 and 13 and numerical sex chromosome abnormalities.

[§] Balanced and unbalanced structural rearrangements and extra markers.

[¶] It was explicitly stated that molecular tests could **not** detect these abnormalities.

TABLE 32 *Obstetricians' unconstrained choice relating to rarer abnormalities*

Abnormality	Obstetricians' choice (% of all normal cases)			
	Molecular test only	Karyotyping only	Both tests	Missing data
Extra marker	41.3	10.1	43.1	5.5
Unbalanced structural	38.5	11.9	41.3	8.3
Balanced structural	42.2	11.9	37.6	8.3

tests would not have placed an extra demand on counselling services, a similar figure to the 64% giving that response in the national general survey.

Midwives' test preferences

Midwives in the West Midlands study region with specific training (or registering an interest in prenatal chromosome analysis by joining the informal regional network) were also surveyed to measure their test preferences (see page 49, 'Midwives'). They were asked to assume that a molecular test could be requested for any of the abnormalities specified in *Table 33*. For cases where a patient was not identified to have a specific risk before testing, they were asked which abnormalities they considered a molecular test should be able to detect.

The preferences of midwives reflected a similar pattern to that reported by obstetricians, although fewer were certain that testing for trisomies 13 and 18 was essential, and a slightly higher proportion indicated that testing for X, Y chromosome abnormalities was essential. Also, more midwives recorded 'don't know' for these chromosome abnormalities, with almost half uncertain about XXX detection.

In terms of likely future testing patterns, given an unconstrained choice, the majority of midwives (71.4%) expressed a preference for the molecular test to act as an adjunct to karyotyping. The same number indicated that they would prefer to give patients the freedom to choose the test themselves.

Finally, most midwives considered that the introduction of molecular tests would not place an additional burden on either counselling services (66.7%) or the way in which results are given to patients (85.7%). However, five midwives (23.8%) did indicate that the availability of a molecular test for Down syndrome might lead to an increase in demand for testing.

Stakeholders' test preferences

Parents

In the prospective trial carried out as part of the project, women were recruited to receive either a molecular test plus karyotyping (intervention group) or to receive just karyotyping (control group). For the intervention group, women's test preferences (and their partners) were recorded in the first questionnaire administered at the time of booking for amniocentesis or the visit for amniocentesis. Those in the control group were not asked about alternatives to the traditional test.

Women in the intervention group were asked, before amniocentesis, to choose between molecular tests and karyotyping. Significantly more women ($p < 0.001$) stated an *ex ante* preference for the molecular test (66.7%) than chose karyotyping (31.9%), and only 1.4% had no

TABLE 33 *Midwives' choices for structure of molecular tests*

Abnormality	Essential (%)	Desirable (%)	Not required (%)	Don't know (%)
Trisomy 18	28.6	38.1	9.5	23.8
Trisomy 13	28.6	38.1	9.5	23.8
XXY (Klinefelter) syndrome	9.5	33.3	19.0	38.1
XYY syndrome	9.5	33.3	19.0	38.1
XXX syndrome	11.8	17.6	23.5	47.1
45X (Turner's) syndrome	14.3	47.6	14.3	23.8

preference. Women were asked the same question again 4 weeks after receiving their karyotyping result. At this time 51.7% of women stated they would choose a molecular test, 42.7% would choose karyotyping, and 5.6% stated they had no preference. Therefore, *ex post,* women continued to state a preference for molecular tests (χ^2, $p < 0.02$), but some women changed their initial preference for a molecular test after the testing process was completed (McNemar test, $p = 0.019$).

When partners were asked to state their preferences at recruitment, the majority (54.4%) stated a preference for the molecular test rather than karyotyping (45.6%), although the difference was not significant ($p = 0.595$). This pattern of responses was similar to that reported by women.

One concern for those examining the costs of introducing molecular tests is that the long wait for a test result may be a factor for patients in deciding whether to undergo amniocentesis. In informal interviews obstetricians and midwives were of the opinion that very few women were deterred from having the test by the comparatively long waiting period for results. Clinicians and patients opted for amniocentesis by weighing the risk of miscarriage against the chances of a fetus with a chromosome abnormality. Once the possibility of such an abnormality was identified by primary assessment of risk factors few declined amniocentesis.

Health commissioners

Health commissioners were also surveyed to explore their preferences, as discussed on page 49, 'Health commissioner surveys'. Seventy-six (72.4%) of those mailed either responded personally or passed the questionnaire to the relevant colleague for completion. Only 12 respondents (15.8%) stated that they would be willing to commission both tests. Otherwise, preferences were evenly divided between those who would consider commissioning molecular tests for women (36.8%) and those who would wish to offer only karyotyping (31.6%). A further 14.5% were unwilling to state a preference. Of the remainder, one respondent wished to offer women the choice and one wanted a test algorithm to be developed and used.

The questionnaire also invited free text comments. Several commissioners reported that they were interested in the use of molecular tests but thought that there was, as yet, insufficient evidence on which to make a judgement. Many indicated they might consider offering karyotyping only for high-risk cases; there seemed little interest in offering both a molecular and traditional test.

General public

A survey similar to that for the health commissioners was used to assess the opinions of the general public. This questionnaire, offering a constrained choice between molecular tests and karyotyping, was sent during May 1999 to 1000 people in one Health Authority in the West Midlands (Solihull). The sample was drawn at random from the health authority register following ethics approval. In total, 289 responses were received; 118 males (40%) and 163 females (56%); the sex of eight respondents was unknown. Analysis of the age of respondents, and their household income, showed that the sample responding was skewed towards older respondents and those with higher incomes.

Six respondents indicated that they would prefer not to be tested for chromosome abnormalities and two indicated they would prefer both tests, although this was not an option on the questionnaire. Of the remainder, 59% stated a preference for karyotyping compared with 38% for the molecular test; the remainder stated an alternative preference. Test choice was independent of age, gender, income or number of children.

Discussion

As can be seen by the responses to the surveys, selection of patients for amniocentesis and prenatal testing is currently largely decided locally, either at a departmental level or by individual clinicians. Only 26% of Trusts reported that they had a formal written protocol and a further 28% reported informal protocols. Therefore, the manner in which molecular tests are likely to be used and, particularly, the extent to which they might replace the existing diagnostic procedure (Fineberg level 3) will depend on the preferences of clinicians, women and other stakeholders. Preferences differ depending on the group being considered:

1. When given a constrained choice of either molecular tests **or** karyotyping, it appears that obstetricians do not prefer karyotyping. Overall, only 9% of UK obstetricians would choose karyotyping; 40% would select a molecular test; 24% would give women the choice; and 27% were undecided.

2. If molecular tests were to replace karyotyping, the option preferred by obstetricians is a molecular test that can detect all common abnormalities. Just over half the obstetricians (57%) reported that in general they would choose a test for the five common abnormalities, although 41% would prefer a more flexible test; only 2% expressed a preference for a Down only test. When asked about specific cases retrospectively, a similar percentage stated a preference for multiplex molecular tests (in 62% of cases). However, for their own patients clinicians were slightly more positive about a Down only molecular test (21% of cases), and only in 17% of cases would they have opted for a more flexible test.
3. When obstetricians were asked to state their preferences given an unconstrained choice, molecular tests were once again the preferred option. Fifty-seven per cent of obstetricians expressed a preference for molecular tests for the majority of their patients and karyotyping for a minority. Only 15% would choose both tests for all, or a majority, of women.
4. Similarly, when women were asked, before karyotyping, to state their preferences, 67% expressed an *ex ante* preference for molecular tests; only 32% would have preferred karyotyping. Once karyotyping test results had been received, however, fewer women expressed a preference for the molecular test (52%), although the change was not statistically significant ($p < 0.02$), and 43% now reported a preference for karyotyping. Partners' preferences before karyotyping were similar, at 54% for molecular testing and 46% for karyotyping.
5. In the scenarios where molecular tests are used as a full or partial replacement for karyotyping it has not been possible to investigate the possible harmful effects of failure to detect, prenatally, a chromosome abnormality. In effect, the results of failure to detect an abnormality can be regarded as giving the parent a false-negative result. When molecular tests are used without karyotyping, within a screening programme, false-negative results will arise from two sources – false-negative serum tests and/or primary risk assessment and abnormalities undetected after amniocentesis. Analysis of the costs and consequences of these false-negatives is outside the scope of this study. Furthermore, there is little good evidence[65,66] on the effects on the parents and their future parenting, on medico-legal issues and on confidence in screening; this is both an important matter for further research and for counselling.
6. Commissioners were evenly divided in terms of their test preferences, with 37% choosing molecular tests and 32% karyotyping. Only 16% reported that they would be willing to commission both tests.
7. In contrast, the general public was unusual in expressing a preference for karyotyping (60%), although the sample of respondents was skewed towards the older and higher income sections of the public.
8. When asked to identify which specific chromosome abnormalities, other than Down syndrome, molecular tests should be able to disclose, half the obstetricians surveyed considered testing for trisomies 13 and 18 essential, but far fewer (6–15%) thought numerical sex chromosome abnormalities essential. The former chromosome abnormalities (trisomies 13 and 18) are known to lead to physical disability and severe mental disability, and have a high mortality in the first few weeks; the latter may lead to milder impairment in intelligence and other mild effects (*Table 1*). Midwives were even less likely to identify the need to test for abnormalities other than Down syndrome; only 29% reported they would test for trisomies 13 and 18 and a similar proportion (10–12%) for numerical sex chromosome abnormalities.
9. Within the West Midlands and Northern Ireland obstetricians were also asked about *ex post* test preferences for specific women. They were given data on rarer abnormalities, which molecular tests are currently unable to disclose, and asked to revisit their preferences for these individuals. In the light of this information, in approximately 10% of cases they continued to state an *ex post* preference for karyotyping only; in 38–42% of cases they would have requested a molecular test only; and in 38–43% of cases clinicians now indicated a preference for offering patients both tests, up from only 15% (see 3 above).
10. Obstetricians were also asked whether they would prefer to test women routinely for these rarer abnormalities. Depending on the abnormality, only 16–22% would test all higher-risk women without exception; 66–70% would only test if a specific risk had been identified in that woman; and 10–12% were unsure.

One of the conclusions that may be drawn from the surveys is that a test which detects clearly defined abnormalities presents fewer ambiguities in terms of consenting and patient management. Thus the higher proportion of 'don't knows' relating to the numerical sex chromosome

abnormalities and other rarer abnormalities may result from a number of factors, including:

- variable and sometimes difficult to predict degrees of handicap associated with a detected abnormality
- less experience and clinical knowledge, as these are rare events in an individual's clinical career.

In such cases it may be more difficult for the clinician to give unequivocal clinical advice when counselling the parents. However, both obstetricians and midwives were of the opinion that the introduction of molecular tests would not lead to an additional burden in terms of counselling requirements.

Chapter 7

Estimating the cost-effectiveness of molecular tests

Introduction

In order to inform the introduction and use of molecular tests, it is necessary to have some method for comparing the costs and benefits of different tests. As discussed, for most diagnostic tests, benefits are usually measured in terms of the accuracy of test results, the impact of results on patient management and, ideally, the impact of diagnostic information on patient outcome. As shown in chapter 2 there was no significant difference in diagnostic accuracy between new molecular tests and existing technologies for the chromosome abnormalities that the molecular test is designed to detect (i.e. absolute sensitivity/specificity). Therefore, the most obvious benefit of molecular tests compared to the traditional test is a faster test result.

One crucial potential disbenefit of diagnostic molecular tests, however, is that rarer abnormalities are not tested for and therefore are more likely to remain undetected, possibly until after the birth of the baby. Thus, there is no direct equivalence between the existing test (karyotyping) and molecular tests. The relative sensitivity and specificity between karyotyping and molecular tests were also considered in chapter 2.

In terms of impact on patient outcome, assessment of diagnostic tests conventionally focuses on patients who receive positive test results. However, only 3.2% of the population tested prenatally for chromosome abnormality will receive a positive (abnormal) result. At the same time, the benefits of speedier results are also experienced by the 96.8% who receive results that indicate that the fetus is normal (that is, no abnormality has been detected).

During the second stage of the study, when molecular tests were trialled in a service setting, the final two levels in the Fineberg assessment hierarchy were addressed:

- **Patient outcome** – does the test contribute to improved health or reduced anxiety for the patient?
- **Cost-effectiveness** – does use of the molecular test improve the cost-effectiveness of healthcare compared to alternative interventions?

During this stage of the study, a group of women undergoing karyotyping and those receiving two tests (a molecular test and karyotyping) were studied. The health status of all women and their partners was measured at the time of testing, and throughout the period of waiting for test results. At the time of testing, women (and their partners) who would receive abnormal results could not be differentiated from those who would receive normal results. Therefore, the sample size that would have been required to adequately record changes in health status for women with a positive or abnormal result precluded the possibility of measuring outcomes for this group of women. Thus, the health benefits of molecular tests were only measured quantitatively for those receiving a negative result.

A further aim of the project was to provide an opportunity for reflection on the main methodological approaches commonly used for comparing costs and benefits – namely cost-effectiveness, cost–utility and cost–benefit analysis. In particular, the study aimed to compare these approaches in the evaluation of diagnostic tests. For the specific case of prenatal testing for chromosome abnormalities, the molecular tests were assessed using the following three approaches:

1. The number and types of abnormalities detected over an extended period of time were analysed in order to enable a cost-effectiveness analysis (calculated as a cost per case detected) to be performed.
2. Health status/QoL was measured during the testing process (using EuroQoL EQ-5D), both for women undergoing karyotyping and for those receiving molecular results, in order to enable a cost–utility analysis to be performed.
3. Contingent valuation (WTP) was used to explore the values placed on more rapid results by women undergoing testing, their partners, health commissioners and the

general public, thus addressing valuation within a cost–benefit framework.

In addition, the cost-effectiveness analysis was further extended by adding a function based on the cost per day of anxiety avoided, weighted by a measure of diagnostic uncertainty.

The remainder of this chapter explores the first of the three approaches above, cost-effectiveness analysis. This is set within the context of different testing regimes that might be implemented locally or nationally. These include the use of molecular tests as a replacement for karyotyping; molecular tests as a routine add-on to karyotyping; and testing regimes with varying levels of karyotyping replacement. Findings based on the other two approaches (cost–utility and cost–benefit analysis) are reported in the next chapter, together with the results of the extended cost-effectiveness analysis.

Cost-effectiveness of molecular tests as a replacement for karyotyping

In order to calculate the cost-effectiveness of a new diagnostic test versus the status quo, a single common outcome measure is required.[67] In the case of a diagnostic test used for medical screening, a commonly quoted comparator is the number of cases detected. This parameter was used in the current study, where cases were defined as all chromosome abnormalities detected by a test without necessarily making any judgement about their clinical significance. An alternative outcome measure used in this area, the number of Down syndrome births avoided,[8] was not considered appropriate for the current assessment. Such a simple outcome measure was rejected because the diagnostic tests being evaluated are all capable of detecting a broader range of chromosome abnormalities than Down syndrome only. Not all of these abnormalities will necessarily result in parents considering termination of the pregnancy, and even if they do this will not necessarily result in a decision to terminate. Furthermore, if the outcome measure used were to focus on termination alone, the benefits of improved information to parents on these other chromosome abnormalities would, by definition, be ignored.

Table 34 shows the cost per case detected for different types of molecular test, if these molecular tests were to be used as a complete replacement for karyotyping for the 37,000* amniotic fluid samples tested per year in the UK (laboratory survey 1998, activity in 1997 and 1998[13]). Two configurations of molecular test are considered:

- 'Molecular all common' – 5-probe FISH or multiplex Q-PCR test configured to detect trisomies 21, 13, 18 and numerical sex chromosome abnormalities on each sample.

TABLE 34 *Cost per case detected for molecular tests as replacement for karyotyping (based on laboratory throughput of 1000 samples per annum)*

Molecular test regime		Direct cost per patient tested	Cost per case[a] detected (% change from status quo)	Per 1000 patients tested	
				Total cases undetected[b]	Clinically significant cases undetected[b]
Down only	FISH	£42.14	£2657 (+23%)	25.5	15.4
	Q-PCR	£44.16	£2784 (+29%)		
All common	FISH	£75.07	£2768 (+28%)	14.3	4.2
	Q-PCR	£47.46	£1750 (–19%)		
Karyotyping (status quo)		£89.68	£2166 (0%)	0	0

[a] *Denominator is number of cases detected by test used*
[b] *Based on audit data from 1988 to 1999 in Birmingham and 1989 to 1999 in Northern Ireland, including cases with no clinical significance*

* Throughput figure reported from UK laboratory survey was 37,780 for 1997. This figure was based on estimates by some respondents as auditing periods varied from service to service. The Advisory Committee on Genetic Testing estimated that 36,817 amniotic fluid samples had been received for testing for 1997–98. Throughout these calculations, it is assumed that 37,000 samples of amniotic fluid are tested per annum in the UK.

- 'Molecular Down only' – FISH or Q-PCR test to detect Down syndrome only on each sample.

Obstetricians' choices for different replacement options for karyotyping were described in chapter 6 (pages 52 and 55). When given a constrained choice in terms of karyotyping replacement, only 1.8% reported that they would choose a Down only molecular test as a replacement; 57.3% of obstetricians expressed a preference for a molecular test for trisomies 21, 13, 18 and X, Y related abnormalities; the remainder (40.9%) reported that they would prefer a test for Down syndrome (trisomy 21) plus the option of selecting from trisomies 13, 18 and X, Y related abnormalities.

Table 34 presents the cost-effectiveness ratios (cost per case detected) for molecular tests and karyotyping. The direct cost per patient tested (in column 1) is based on a laboratory throughput of 1000 samples per annum (close to the UK laboratory median throughput).

These cost figures, like all the costs presented in the following sections, only relate to the direct cost of performing the various diagnostic tests. Other costs, such as those of pre-diagnostic screening tests, prenatal test counselling, amniocentesis, etc. were excluded from the cost analysis because they are expected to be independent of the type of diagnostic test performed. Costs such as sample transport, laboratory reception, patient records management, quality assurance and dispatch of results were also excluded for similar reasons. In addition, the costs of consequences such as medico-legal costs resulting from changes in test regimens have not been considered in this analysis. This means that the cost figures presented in *Table 34* do not represent the full cost of testing, although the cost differences will be representative of actual differences.

Furthermore, the costing method used in the study will tend to lead to a low cost estimate because the staff costs used are likely to be lower than those in many UK laboratories, because these costs were measured at a large laboratory with an optimal skill-mix. As pointed out in chapter 4, in smaller laboratories senior grades may be performing routine tasks that could otherwise be optimised to lower staff grades. This is exemplified by the fact that the mean price for karyotyping reported by UK laboratories was £153 (28/30 responses), whereas in larger, more efficient laboratories it was £120.

Cost-effectiveness ratios presented in *Table 34* (column 2) are based on the prevalence of chromosome abnormalities detected in the two study sites over a period of 10–11 years. The table also provides an estimate of the total number of cases that would not be disclosed per 1000 patients tested (column 3) if molecular tests were to replace karyotyping, and the number of these that would be clinically significant (column 4). Both columns 3 and 4 are based on data on chromosome abnormalities detected in the two study sites, and therefore should be treated with some caution because these sites may not be representative of other sites nationally. Also, even for the clinically significant cases that would not be detected, it should be remembered that these chromosome abnormalities will have varying levels of impact on the child (see page 16, 'Fineberg level 2: molecular test diagnostic accuracy). These may range from relatively minor effects to those that probably mean that the fetus will not survive to birth (*Table 1*). Furthermore, around half of cases of Down syndrome (51% per annum) are still not identified prenatally, but diagnosed clinically soon after birth.[68]

The cost-effectiveness ratios presented in *Table 34* indicate that the current gold standard (karyotyping) is more cost-effective than all options but multiplex Q-PCR. This apparent contradiction of cheaper molecular tests but lower cost-effectiveness arises because the molecular tests detect fewer abnormalities than karyotyping, although karyotyping is the most expensive test. Changing the configuration of a molecular test to detect a wider range of abnormalities is found to increase the cost per positive case detected for FISH by 5%, but the same ratio for Q-PCR is decreased by 48%. This is because for Q-PCR the marginal cost required to extend the test is more than counterbalanced by the increased number of cases detected, whereas for FISH this is not the case.

As explained earlier, the cost-effectiveness ratios given in *Table 34* are based on a throughput of 1000 samples per annum. However, Q-PCR is highly sensitive to throughput (see page 37, 'Cost per sample tested', and *Figure 8*), with the cost per test rising significantly with lower throughputs. Because of this, karyotyping is calculated to be a more cost-effective option in laboratories with throughputs below approximately 750 samples per annum. Once throughput reaches 1500 samples per annum, the cost-effectiveness ratio for multiplex Q-PCR decreases from £1750 to £1393 per case detected, a ratio that is now 36% less than that for karyotyping at

the same throughput. Q-PCR tests will therefore become more cost-effective than karyotyping under the following conditions: multiplex for laboratories with volumes higher than 750 samples and Down only configuration for laboratories with a throughput higher than 1300 samples per annum.

As well as throughput, certain assumptions made in the costing method may also have an effect on the relative cost-effectiveness ratios for different tests. For example, basing test costs on a large laboratory with appropriate skill-mix will produce a relatively more favourable ratio for karyotyping (and to some extent FISH) where a large cost element consists of staffing costs. Similarly, the assumption of a dedicated DNA analyser in the cost analysis will maximise capital equipment costs for Q-PCR and this will, once again, favour karyotyping and FISH to a certain extent, by not making allowance for the fact that equipment may be shared in some laboratories.

Overall, therefore, the cost-effectiveness ratios in *Table 34* may be slightly more unfavourable (higher) for Q-PCR tests and somewhat more favourable (lower) for FISH and karyotyping than would be the norm in a UK laboratory with a throughput of 1000 samples per annum.

Finally, for molecular tests the 'cases detected' figure used to calculate the cost-effectiveness ratio includes only clinically significant abnormalities (trisomies 21, 13, 18 and numerical sex chromosome abnormalities). However, the same figure for karyotyping includes **all** abnormalities detected by the test, some of which may be considered clinically of little or no significance. If all non-clinically significant cases were excluded, the cost per case detected for karyotyping would increase from £2166 to £2865. Thus, comparable cost-effectiveness ratios, based on clinically significant cases only, would make all four molecular tests appear more favourable than karyotyping, even at a throughput of 1000 samples per annum.

Cost-effectiveness of molecular tests as a routine add-on to karyotyping

Table 35 presents the calculated cost per case detected for a second testing regime in which all women are offered a molecular test for common abnormalities in addition to karyotyping. Thus, a 'new gold standard' can be defined that offers additional benefits in terms of a rapid first result without loss of diagnostic information. For this 'new gold standard' test regime, amniotic fluid or chorionic villus samples[†] can once again be tested using one of two test combinations:

- 'New gold all' – karyotyping plus a molecular 5-probe FISH or multiplex Q-PCR test (to detect trisomies 21, 13, 18 and numerical sex chromosome abnormalities).
- 'New gold Down only' – karyotyping plus a FISH or Q-PCR test (to detect Down syndrome only).

In the national survey obstetricians were asked to indicate their preferences if, in addition to Down syndrome (trisomy 21), a molecular test could be configured to include any of the specific abnormalities listed in the 'new gold all' option above. Approximately half the clinicians (50–53%) considered a test for trisomies 13 and 18 would be essential, and fewer than one in six considered a test for X, Y chromosome abnormalities was essential (6–15%).

TABLE 35 *Cost per case detected for molecular tests as an add-on to karyotyping ('new gold standard' based on a throughput of 1000 samples per annum)*

'New gold standard' test regime		Direct cost per patient tested	Cost per case[a] detected (% change from status quo)	Total cases undetected per 1000 patients tested
Down only	FISH	£131.82	£3183 (+47%)	0
	Q-PCR	£133.84	£3232 (+49%)	
All common	FISH	£164.75	£3978 (+84%)	0
	Q-PCR	£137.14	£3312 (+53%)	
Karyotyping (status quo)		£89.68	£2166 (0%)	0

[a] *All cases detected by test, including clinically non-significant cases for karyotyping*

[†] The assumption is made that an adequate volume of liquor is available for both processes for all samples.

Table 35 shows the cost per case detected if both karyotyping and a molecular test were performed for all samples, compared to the status quo of karyotyping alone.

Clearly, offering the 'new gold standard' to all women is a less cost-effective option than the status quo, but arguably it will provide the most complete information and reduction from anxiety, assuming all those tested know of the risk of chromosome abnormalities other than Down syndrome. For this test regime, the most cost-effective molecular test (Down only FISH), is still 47% more expensive than karyotyping in terms of the cost per case detected.

Once again, the values given for Q-PCR in *Table 35* are particularly sensitive to throughput, and will change relative to the other tests as laboratory workload rises. At throughputs above 1100 samples per annum, Down only Q-PCR is the most cost-effective of the molecular tests to use in the 'new gold standard' test regime. If the 'all common' test is used, Q-PCR becomes the more cost-effective molecular test to use above 1200 samples per annum. However, even in a UK laboratory operating at the highest UK throughput recorded (5000 prenatal samples per annum, 1998 figure), the most cost-effective molecular test at this throughput (Down only Q-PCR) will still cost 24% more per positive case detected than the status quo.

If the cost-effectiveness analysis is limited to detection of clinically significant cases only, the cost per case detected for karyotyping rises to £2868 per case detected. At a throughput of 1000 samples per annum, although all 'new gold standard' regimes would still be less cost-effective than karyotyping, the cost per case detected using Down only FISH as an add-on to karyotyping, to give a more rapid result for Down syndrome, would be only 11% higher than for karyotyping alone.

Cost-effectiveness of testing regimes with varying levels of karyotyping replacement

Clearly the service level impact on costs and benefits of the introduction of molecular tests into laboratories will depend on the actual approach to their introduction that is adopted. This may be more complex than the approaches discussed above. For example, the surveys of obstetricians and midwives (pages 55 and 57) indicated that for some patients karyotyping will always be considered the most appropriate test. In addition, clinicians identified a small subgroup of patients for whom it may be desirable to offer both a molecular test and karyotyping. It is important, therefore, to explore the costs and cost-effectiveness of a range of testing regimes that fall somewhere between the two extremes discussed in the previous sections, that is molecular tests as a replacement for karyotyping and molecular tests as a routine add-on to karyotyping.

Table 36 summarises the five possible testing regimes suggested by respondents to the surveys, including the two discussed above. The remainder of this chapter first presents estimates of the marginal cost of each testing regime at a UK level, based on reported annual activity figures. Then, where possible, these costs are compared to the number of abnormalities that would be disclosed by the test configuration used, in order to calculate a marginal cost per case detected.

TABLE 36 *Possible testing regimes arising from the national obstetricians' survey*

Testing regime		Summary	
1		'New gold standard', molecular test and karyotyping for all samples	
2	A	Local or national decision: 'one for all' option	Karyotyping only for all samples
	B		Molecular test only for all samples
3		Molecular test for all women, some samples karyotyped in addition	
4		Karyotyping for all women, selected samples given additional molecular test	
5	A	'Either–or' option	Some samples karyotyping, some molecular test, clinicians' choice
	B		Variant of 5A, based on parental choice

Costs of alternative testing regimes

As in previous sections, the costs presented below relate only to the direct cost of performing the various diagnostic tests. The cost figures will therefore underestimate full test costs. However, as the costs excluded are common to both molecular and non-molecular tests, the cost differences should reflect the true marginal savings (or expenditure) arising from the introduction of alternative testing regimes.

Testing regime 1: 'new gold standard'

The first testing regime, the 'new gold standard', which would provide karyotyping and molecular testing for all samples, has already been discussed on page 64 in terms of its cost-effectiveness ratios. A total UK direct cost for this option can also be estimated, using the cost per test from the costing model and a total annual UK throughput of around 37,000 (see footnote on page 62) women tested per annum. *Table 37* shows that this would result in an additional maximum cost of £2.78 million per annum, compared to current direct costs, if FISH were to be used to detect all common abnormalities, and a minimum additional cost of £1.56 million if FISH were used to detect Down syndrome only. The cost of using Q-PCR would lie between these two, but closer to the latter. As reported in chapter 6, when asked for their likely pattern of testing if given a free choice, only 15% of obstetricians said they would choose both tests for all, or the majority of, women (see *Table 28*).

Testing regime 2: one test for all

In this strategy, a decision would be made locally or nationally to either continue to offer karyotyping to all women and not use molecular tests (option A) or instead only to offer a molecular test to all women (option B). The regime, 'one-for-all-option A', would represent maintenance of the status quo, and 'one-for-all-option B' would represent complete test replacement.

When asked for their likely pattern of testing if constrained to ordering either a molecular test or karyotyping, only 8.8% of obstetricians expressed a preference for option A (i.e. continuing to restrict all testing to karyotyping; see *Table 29*). *Table 38* provides estimates of the total UK direct costs if the 'one test for all' testing regimes were adopted nationally.

The figures indicate that option B, complete replacement of karyotyping, could potentially lead to savings of approximately £0.54–1.76 million per annum if FISH tests were used (5-probe and Down only respectively). Savings of between £1.56 and £1.69 million per annum could potentially be made if Q-PCR were introduced (multiplex or Down only respectively).

A 'molecular test for all' regime could therefore provide the opportunity to generate cost savings of up to 53% of the current cost of test performance (excluding non-direct costs). However, as was evident in *Table 34*, the cost per case detected would be higher for all molecular tests except multiplex Q-PCR.

Testing regime 3: molecular test for all with selected samples also karyotyped

In this approach, a molecular test would replace karyotyping for all samples, but some women would also be offered karyotyping. *Figure 14* shows the impact on cost per person tested if an increasing proportion of women were to be offered karyotyping in addition to a molecular test. The values presented in this figure are based on a laboratory with a throughput of 1000 cases per annum.

The figure illustrates that in this testing regime the cost per person tested is naturally lower than

TABLE 37 *Estimated annual UK test costs (% change from current practice) for molecular tests as an add-on to traditional test*

Abnormalities detected	'New gold standard' molecular technology (£[a], %[b])		Karyotyping (£[a], %[b])
	FISH	Q-PCR	
Down syndrome only	£4.88 m (+47%)	£4.95 m (+49%)	
Trisomies 21, 18, 13 and X,Y abnormalities ('all common')	£6.10 m (+84%)	£5.07 m (+53%)	
Those currently detected			£3.32 m (0%)

[a] *Cost figure only relates to direct test performance (excludes sample transport, lab reception, despatch of results, etc.)*
[b] *% change from direct cost of karyotyping*

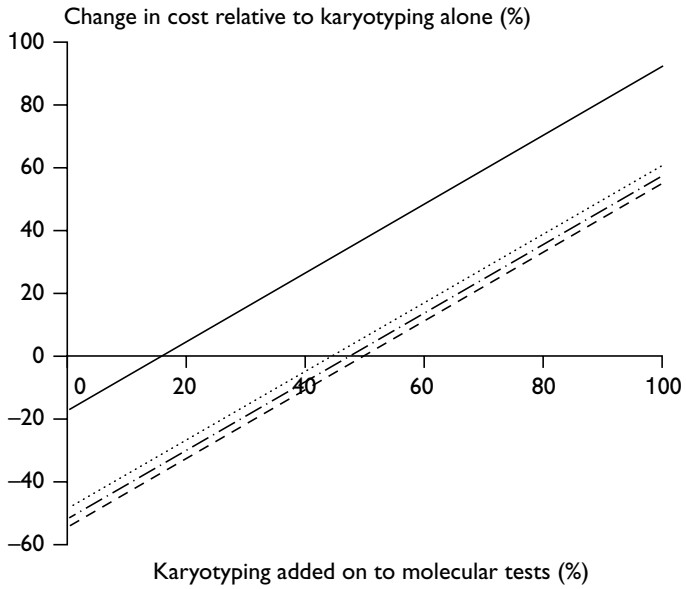


FIGURE 14 *Karyotyping as an add-on to molecular tests for an increasing proportion of women for a laboratory throughput of 1000 samples per annum (——, FISH 5-probe; ·····, Q-PCR multiplex; — · — ·, Q-PCR Down only; – – –, FISH Down only)*

would have been the case if all women were offered karyotyping, as long as the proportion offered both tests is limited. The break-even point for dual testing differs depending on the molecular test being considered.

If the FISH Down syndrome only test were to be used, 53% of women could be offered karyotyping as well as the molecular test for the same cost as karyotyping all women. If the FISH 5-probe test were used, only 16% of women could be offered karyotyping for a neutral cost. For Q-PCR the difference between the cost of the Down syndrome test and the multiplex test is small, and 51% of women could be offered karyotyping as well as a Q-PCR Down syndrome only test, and 47% offered karyotyping plus a multiplex Q-PCR test for a neutral cost. For laboratories with throughputs other than 1000 cases per annum, the impact on cost per person tested is presented in appendix 5, *Table 59*.

The cost per case detected will of course vary depending on the criteria used for selecting women who are to receive both tests. If it is assumed that only high-risk cases would be offered both tests, the data recorded in the West Midlands over a period of 11 years can be used to estimate the number and types of abnormalities that would be detected in both

TABLE 38 *Estimated annual UK test cost differences (% change from current practice) for molecular tests as replacement for traditional test*

Abnormalities detected	Option B: Molecular tests for all (£[a], %[b])		Option A: Karyotyping for all
	FISH	Q-PCR	(£[a], %[b])
Down syndrome only	£1.56 m (–53%)	£1.63 m (–51%)	
Trisomies 21, 18, 13 and X,Y abnormalities ('all common')	£2.78 m (–16%)	£1.76 m (–47%)	
Those currently detected			£3.32 m (0%)

[a] *Cost figure only relates to direct test performance (excludes sample transport, lab reception, despatch of results, etc.)*
[b] *% change from direct cost of karyotyping*

high-risk and low-risk groups (see *Table 10*). For these calculations the mean prevalence of chromosome abnormalities for Northern Ireland and Birmingham was used, as this figure allowed for differences in case selection and prevalence and was seen as more representative than data from one area of the UK. The mean prevalence was applied to 37,000 tests, the estimated UK tests per annum. This figure, although highly sensitive to case selection, enabled cost-effectiveness ratios to be estimated as shown in *Table 39*. These are comparable with the earlier cost-effectiveness ratios presented in *Table 34*, which similarly indicated that the current gold standard (karyotyping) is more cost-effective than all options but multiplex Q-PCR.

Once again, if all non-clinically significant cases are excluded, then the cost per case detected for karyotyping would increase from £2166 to £2865. Thus, cost-effectiveness ratios based on clinically significant cases only indicate that all four molecular tests are more favourable than karyotyping for testing regime 3, even at a throughput of 1000 samples per annum.

Similar information detailing the breakdown of abnormalities in different subgroups was not available for other patterns of dual testing. When presented with a free choice, however, only 8.5% of obstetricians said they would choose both tests for all women, not very different from the 6.6% of high-risk cases shown in *Table 39*.

Testing regime 4: karyotyping for all with additional molecular test for selected patients

In this testing regime, karyotyping would continue to be the test conducted on all samples, but in some cases it may be judged desirable also to offer women a molecular test. *Figure 15* examines the cost implications of such an approach and shows the cost per person tested as the proportion of women offered a molecular test in addition to karyotyping increases. Results are presented for both Down syndrome only tests and for 5-probe FISH and multiplex Q-PCR tests in a laboratory with a total throughput of 1000 samples per annum.

The figure demonstrates that were one in five women to be offered a molecular test in addition to karyotyping, the percentage increase in the cost of testing would be between 9% and 17%. If six out of ten of women were offered a molecular test in addition to karyotyping, costs would increase by 28–50%, depending on the molecular test used.

Figure 15 also illustrates that FISH (Down syndrome only) is the least expensive option at all levels. For Q-PCR there is little to be gained by restricting the molecular test to Down syndrome only as the multiplex test costs little more. In contrast, use of the FISH 5-probe would prove to be a more expensive option than use of the other three tests and arguably will be unnecessary if karyotyping is to be undertaken as well.

The values presented in *Figure 15* are once again based on a laboratory with a throughput of 1000 cases per annum. The cost of offering molecular tests in laboratories with other throughputs is presented in appendix 5, *Table 60*.

Unfortunately, once again it is not possible to estimate accurately the number of cases detected for the various test mixes shown because this will be dependent on the selection criteria used by clinicians. Furthermore, it is impossible to predict

TABLE 39 *Cost per case detected for testing regime 3 (molecular tests for all women), karyotyping for high-risk cases (7%) at throughput of 1000 samples per annum*

Testing regime 3		Direct cost per patient tested	Total UK direct testing cost[a] (% change[b])	Cost per case detected (% change from status quo)	Per 1000 patients tested[b]	
					Total cases undetected	Clinically significant cases undetected
Down only	FISH	£48.06	£1.78 m (–46%)	£1889 (–13%)	15.9	9.1
	Q-PCR	£50.08	£1.85 m (–44%)	£1968 (–9%)		
All common	FISH	£80.99	£2.99 m (–10%)	£2527 (+17%)	9.4	2.5
	Q-PCR	£53.38	£1.98 m (–40%)	£1665 (–23%)		
Karyotyping (status quo)		£89.68	£3.32 m (0%)	£2166	0	0

[a] *Based on 37,000 amniotic fluid samples per annum, see footnote page 62*
[b] *% change from direct cost of karyotyping*
[c] *Based on audit data from 1988 to 1999 in Birmingham and 1989 to 1999 in Northern Ireland*

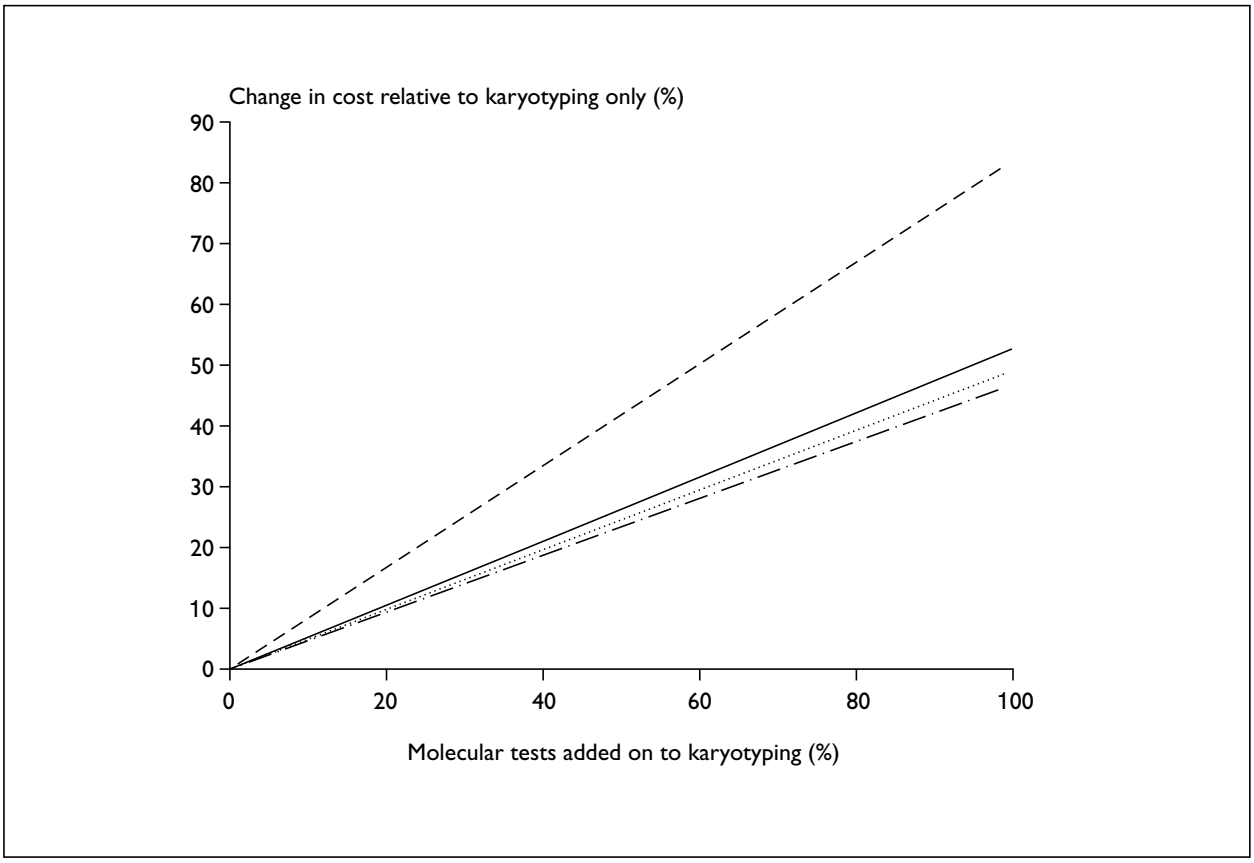


FIGURE 15 *Molecular tests as an add-on to karyotyping for an increasing proportion of women for laboratory throughput of 1000 samples per annum (– – –, FISH 5-probe; ——, Q-PCR multiplex; ·····, Q-PCR Down only; — · — ·, FISH Down only)*

the likely pattern of use of molecular tests in addition to karyotyping if this form of testing regime were to be introduced. However, if it is assumed once again that only high-risk cases would be offered both tests, it is possible to calculate a cost-effectiveness ratio for each of the molecular tests. These values are shown in *Table 40*, which indicates that testing regime 4 would have a minimal impact on UK direct testing costs, and also that cost-effectiveness ratios are largely similar for all four molecular tests and karyotyping.

Testing regime 5A: 'either–or' option

Finally, clinicians suggested that it may be possible to introduce a selective 'either–or' policy. Instead of offering both tests to some or all women, this would mean that some women would be offered a molecular test as the sole test, and the remainder would be offered karyotyping. *Figure 16* shows the impact on test costs of gradually increasing the proportion of karyotyping in the test mix from 0% to 100% (where 100% would be all women given karyotyping, i.e. the status quo).

TABLE 40 *Cost per case detected for testing regime 4 (karyotyping for all women, molecular test for high-risk cases) based on a throughput of 1000 samples per annum*

Test regime 4		Direct cost per patient tested	UK direct testing cost[a] (% from status quo)	Cost per case detected (% change from status quo)	Total cases undetected per 1000 patients tested
Down only	FISH	£92.46	£3.42 m (+3%)	£2233 (+3%)	0
	Q-PCR	£92.59	£3.43 m (+3%)	£2236 (+3%)	
All common	FISH	£94.63	£3.50 m (+6%)	£2285 (+6%)	0
	Q-PCR	£92.81	£3.43 m (+4%)	£2241 (+3%)	
Karyotyping (status quo)		£89.68	£3.32 m	£2166	0

[a] *Assumed 37,000 tests per annum in the UK, see footnote page 62*

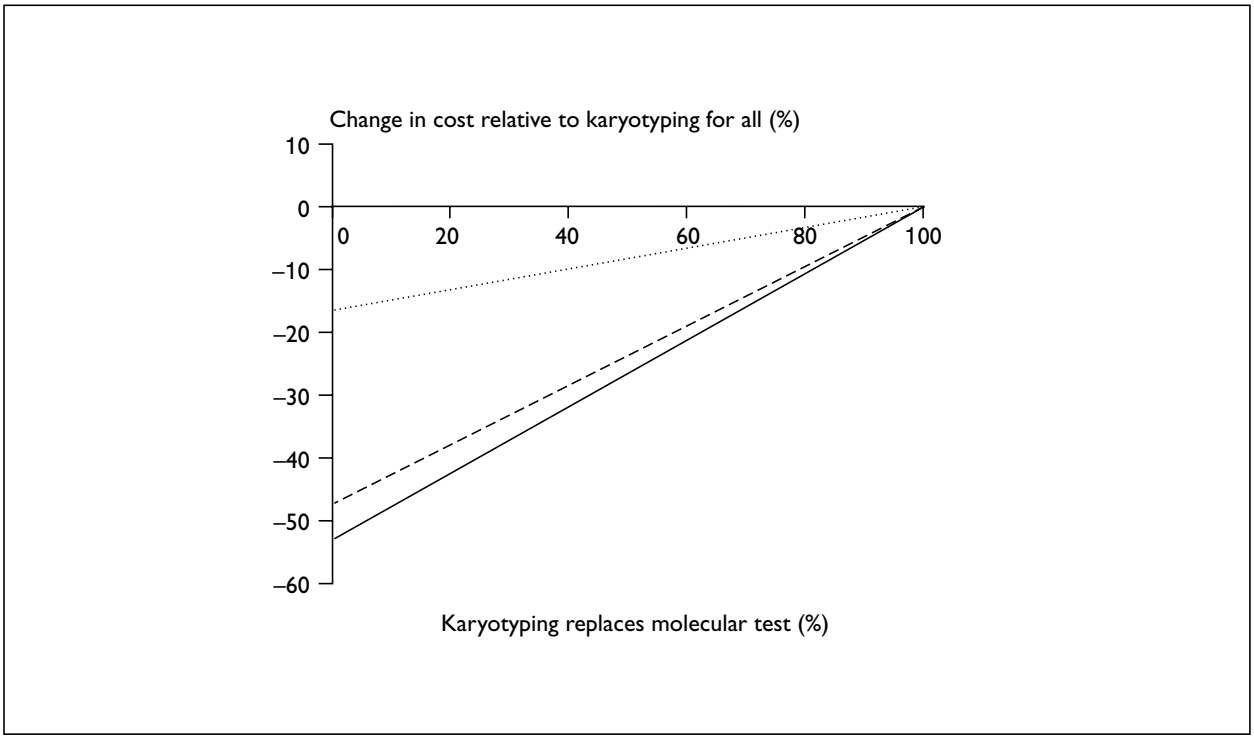


FIGURE 16 *Karyotyping as a replacement for molecular tests for an increasing proportion of women for laboratory throughput of 1000 samples per annum (·····, FISH 5-probe; – – – –, Q-PCR multiplex; ———, FISH Down only)*

For the sake of clarity the trend line for Q-PCR Down syndrome only has been omitted because it falls exactly halfway between FISH Down syndrome only and multiplex Q-PCR. Once again, the results shown in *Figure 16* are based on a laboratory throughput of 1000 cases per annum and data for other throughputs are given in appendix 5, *Table 61.*

Figure 16 illustrates that for the 'either–or' test regime, all combinations of molecular and karyotyping tests will produce savings on the status quo. Also, use of multiplex Q-PCR as the molecular test will prove to be the least expensive option, even in laboratories with throughputs as low as 500 cases per annum. Furthermore, for larger laboratories with throughputs greater than 1600 cases per annum, all women could be offered a multiplex Q-PCR test for the same cost as a FISH Down syndrome only test that detects fewer abnormalities.

As reported in chapter 6, however, when clinicians were asked for their likely preferences if constrained to ordering either a molecular test or karyotyping, 8.8% expressed a preference for restricting testing to karyotyping; a further 27.3% reported 'don't know'; and 23.9% said the woman should choose (see *Table 29*).

Testing regime 5B: parental choice

Chapter 6 reported parental preferences for the type of test performed when a constrained choice is made between either a molecular test or karyotyping. One variation of the 'either–or' testing regime is that parents be given the 'either–or' choice. It can be assumed that all high-risk cases would, in this regime, be given either karyotyping or a 'new gold standard' test. Approximately 7% would be likely to be offered karyotyping or the 'new gold standard' as they are in the high-risk group and of the remainder, 60% of women in the study indicated that they would request a molecular test *ex ante* and 33% could be assumed to opt for karyotyping. The cost-effectiveness ratio for the various test configurations are shown in *Table 41,* together with numbers of clinically significant cases not detected using this test regime.

If parental choice is used to guide test selection, savings can be made over the current regime. So long as all high-risk cases are offered a test configuration that includes karyotyping then fewer abnormalities are missed than test regime 3, although more moderate cost savings will also be made. Thus, while a majority of parents request molecular tests alone, savings can be made on all test configurations except 5-probe FISH and

TABLE 41 *Direct test costs and cost-effectiveness ratios when parents asked to choose either karyotyping or a molecular test*

Test regime 5B		UK direct testing cost[a] (% change from status quo)	Cost per case detected (% change from status quo)	Total cases undetected[a]	Cost per clinically significant case detected (% change from status quo)	Clinically significant cases undetected[a]
Down only	FISH	£2.37 m (–28%)	£2014 (–7%)	9.6	£2480 (–14%)	5.4
	Q-PCR	£2.42 m (–27%)	£2057 (–5%)		£2532 (–12%)	
All common	FISH	£3.18 m (–4%)	£2408 (+11%)	5.6	£2890 (+1%)	1.5
	Q-PCR	£2.50 m (–25%)	£1891 (–13%)		£2270 (–21%)	
Karyotyping (status quo)		£3.32 m	£2166 (0%)	0	£2865 (0%)	0

[a] *Per annum per 1000 women tested*

compared to test regime 3 (karyotyping for high-risk cases only), fewer abnormalities will be missed for very little extra spending.

If parental opinion were to swing the other way with, for example, 30% requesting molecular tests, cost per case detected for multiplex Q-PCR would rise by 24%, to 3% more than karyotyping. If calculations include only clinically significant cases, then the cost per case detected would rise by 13% but savings of 8% would be made on costs. However, 27 clinically significant cases per 1000 women tested would remain undetected for this testing regime.

Incremental cost-effectiveness ratios

The differences between the choices that can be made between test configurations and between and within test regimes can be expressed in terms of incremental cost-effective ratios. These ratios are calculated by dividing the cost differences (C_a–C_b) by the difference in effectiveness (E_a–E_b) for the two options, a and b, under consideration. The measure of effectiveness used for this calculation is the number of cases detected.

Table 42 compares the incremental cost-effectiveness ratios with current practice for various test regimes and configurations for both FISH and Q-PCR. As can be seen, when molecular tests are used as a replacement for karyotyping (test regime 1), the most cost-effective option is multiplex Q-PCR. This is because the gains in effectiveness in terms of cases detected for the multiplex test far outweighs the extra cost of the test for Q-PCR, whereas the cost difference between FISH 5-probe and karyotyping is much less for the same effectiveness.

Table 43 shows that if FISH is used for the molecular test of choice the difference in using a test configured to detect all common abnormalities as opposed to a test configured to detect Down syndrome only is an expensive option. If Q-PCR is chosen as the molecular test, the difference between using a multiplex Q-PCR configuration

TABLE 42 *Incremental cost ratios for various test regimes and configurations compared to current practice of karyotyping for all women*

Test scenario		Molecular test used	Test configuration	
			Down only	Trisomies 21, 13, 18 and X,Y chromosome abnormalities
1	Molecular tests as a replacement for karyotyping for all women	FISH	£68,980	£66,050
		Q-PCR	£37,800	£109,240
3	Molecular tests for all women, karyotyping in addition for high-risk cases	FISH	£96,850	£92,150
		Q-PCR	£34,210	£142,890
5B	Parental choice with gold standard for high-risk cases	FISH	£98,680	£23,130
		Q-PCR	£93,460	£145,000

TABLE 43 *Incremental cost-effectiveness ratios comparing configuring molecular test to detect trisomies 21, 13 and 18 and sex chromosome abnormalities (FISH 5-probe, Q-PCR multiplex) versus Down syndrome only configuration*

Test scenario		FISH	Q-PCR
1	Molecular test as a replacement for karyotyping	£108,787	£10,902
3	Molecular for all, karyotyping for high-risk cases	£187,448	£18,785
5B	Parental choice with gold standard for high-risk cases	£205,472	£20,591

and a Down only test configuration is much less, indicating that it may be cost-effective to use a multiplex Q-PCR test in all situations.

Discussion

Adopting a simple cost-effectiveness approach, it is possible to estimate the cost per case detected for the four molecular test variants, and compare these figures with the value for karyotyping.

The results demonstrate that, if obstetricians and scientists were to agree that a molecular test and karyotyping should be carried out on all samples ('new gold standard'), then in a laboratory with a throughput of 1000 samples per annum the most cost-effective molecular test to use would be a Down only FISH test. However, use of this test regime would increase the cost per case detected by 47% (*Table 35*), at an estimated increase in UK annual testing costs of £1.56 million (*Table 37*). Use of other molecular tests as an add-on to karyotyping will increase the direct test cost per case detected even further at this throughput. However, even if testing were to be conducted in larger laboratories, for example those with throughputs of 5000 samples per annum, the 'new gold standard' test regime would still be approximately one-quarter more expensive per case detected than karyotyping (27%). Therefore, for detection of all clinically significant cases karyotyping on its own is judged to be a more cost-effective option than its use in conjunction with a molecular test, the 'new gold standard'. However, using two tests might be considered worthwhile because such a test combination would reduce anxiety during the waiting period by giving some information sooner, albeit at a greater cost.

For karyotyping alone, even if the denominator in the calculation of cost per case detected is limited to clinically significant cases only, then the cost-effectiveness ratios still indicate that karyotyping on its own is the most cost-effective test at a throughput of 1000 samples per annum.

The introduction of a 'new gold standard' test regime would inevitably increase annual UK testing, by an estimated £1.56 million to £2.76 million per year at 1998 prices (*Table 37*). This estimate is based on test costs in laboratories with a mean throughput of 1000 samples per annum, although there are some economies of scale to be gained, particularly for Q-PCR tests. The costs quoted here are direct costs only. For the 'new gold standard' option indirect costs would also rise as two sets of results would need to be sent to patients. Additionally, in a few cases a molecular test reporting that no abnormality has been detected could be followed by an abnormality detected by karyotyping. Consenting and counselling in this situation thus would be more complex than at present for karyotyping alone.

The 'new gold standard' test regime does not appear to be widely supported by clinicians. When given a free choice, only 15% of obstetricians said they would choose both tests for all, or a majority, of women.

If, instead of molecular tests being used as an add-on test, they were to replace karyotyping, this would naturally reduce the cost of prenatal diagnostic testing in the UK. At a national level this could result in savings of between £0.54 million and £1.76 million per annum, depending on the test introduced (*Table 38*). However, except for multiplex Q-PCR, all molecular tests used as a replacement are calculated to be less cost-effective than karyotyping at a mean throughput of 1000 samples per laboratory (*Table 34*). This is because the benefit of any reduced cost for molecular tests will be outweighed by a reduction in the number of cases detected. This ratio will vary to some extent, depending on laboratory throughput. For example, at throughputs below 700 samples per annum, karyotyping will remain the most cost-effective option. However, at a throughput of 1500, Q-PCR will exhibit a substantially lower cost per case detected (–36%) than karyotyping. The cost-effectiveness ratio for karyotyping includes all abnormalities detected. If non-

clinically significant cases are excluded, then all four molecular tests are judged to be more cost-effective even at a throughput of 1000 samples per annum.

A third testing regime that might be introduced is one in which a molecular test would be performed for everyone, but a karyotyping test would also be carried out for all high-risk women (about 7% of cases). In this testing regime, based on the cost per case detected, multiplex Q-PCR would be the most cost-effective molecular test to use and 5-probe FISH the least cost-effective (*Table 39*).

In any testing regime in which molecular tests replace some or all karyotyping tests, some chromosome abnormalities will remain undetected. The number will depend on the type of molecular test used and case selection for amniocentesis. If a molecular test capable of detecting all common abnormalities were to be used, it is estimated that each year in the UK approximately 2.5–4.2 cases of clinically significant abnormalities per annum per 1000 women tested would not be disclosed. This figure would fall to 1.5 cases of clinically significant abnormality per annum per 1000 women tested if all high-risk women were to be offered karyotyping and a molecular test and the remainder were asked to choose a test (*Table 41*). In this case, the chromosome abnormalities not detected will not be Down syndrome, and not all would result in a decision to terminate the pregnancy. Among these cases, possibly half may be mosaicisms, for which there is a degree of uncertainty (see page 17, 'Clinical significance'). Currently, around half the cases of Down syndrome are not identified prenatally, but are diagnosed clinically soon after birth (National Down Syndrome Cytogenetic Register: Spring 2001).

Finally, as has already been discussed, the costs and cost-effectiveness of particular testing regimes will be dependent on laboratory throughput. In locations where laboratories are smaller, then the most cost-effective test to use in most situations would be FISH, and in larger laboratories Q-PCR. However, chapter 3 also highlighted that a laboratory's work may not consist solely of prenatal testing, and therefore other activities can legitimately influence a laboratory's decision on which test to introduce.

In the current climate, the use of prenatal testing is determined by individual clinicians, laboratories and hospitals. The national survey demonstrated a lack of equity of provision, and evidence of regional and local variations in practice with regard to primary risk assessment. This may well be replicated with regard to prenatal diagnosis if molecular tests were to be introduced without discussion of appropriate implementation protocols. Crucial to this debate will be the needs and wishes of parents and other stakeholders, as well as the views of obstetricians and midwives. These are considered in more detail in the next chapter.

Chapter 8

Valuing benefits for parents and other stakeholders

Introduction

The final level of the adapted Fineberg evaluation hierarchy focuses on the 'cost-effectiveness' of a diagnostic test. This is a generic term, one that may include assessment of impact on health status or QoL of patients, as well as estimates of the value placed on different interventions or strength of preferences for competing alternatives. Any judgement about the value of different prenatal testing regimes cannot be informed solely by comparison of simple cost-effectiveness ratios. Judgement should be extended to include consideration of test preferences for diverse stakeholders and impact on QoL. The economic techniques that lend themselves to measuring these parameters – cost–benefit analysis and cost–utility analysis – were each addressed in the study.

In order to assess molecular tests in these ways, both types of molecular test were trialled in a service setting with women being offered either karyotyping alone or a molecular test and karyotyping. Recruitment methods and patient populations were described in chapter 5 (page 45). Survey instruments were used to obtain QoL measures and test preferences from women as reported below.

Currently, when making a choice about prenatal diagnosis following screening, the main decision that women have to make is whether or not to undergo the invasive procedure required to obtain a sample for testing. Once this choice is made, only one diagnostic test is available.* There is evidence to demonstrate that wide variation exists in information giving and patient understanding about the range of results possible from this test (karyotyping). Thus, women's preferences for new prenatal diagnostic tests giving alternative ranges of results had to be measured in a context in which patient information (and hence understanding) about the current test may be incomplete. Preferences were also recorded for other stakeholder groups (clinicians, health commissioners and the general public).

Parental anxiety

Choice of instrument

In cost–utility studies, it is common practice to estimate the effects of healthcare interventions on health status using a generic instrument to measure QoL. However, in the case of diagnosis in general, and specifically for prenatal testing, anxiety can also play a major role in the patient's health state. Thus, an anxiety-specific instrument was used in the prospective study, in addition to a generic QoL measure. The instrument used, the Speilberger State–Trait Anxiety Inventory, enables quantification of the level of anxiety experienced at any point in time (State), while allowing for the inherent anxiety (Trait) normally felt by the subject. Although the generic instrument used (EuroQoL EQ-5D) had a dimension relating to anxiety, it was anticipated that this dimension would not be as sensitive to changes in a patient's anxiety states as a more specific outcome instrument.

Anxiety levels recorded

Figure 17 presents a graphical display of Speilberger Anxiety State scores recorded for the molecular test group and the control group at various points in time. For women in the intervention group (solid line), these points included (i) at booking or clinic visit; (ii) at molecular test result; (iii) at result of karyotyping; and (iv) 4 weeks after karyotyping result (*Table 22*). The intervention group were offered the 'new gold standard' test regime.† For the control group

* CVS and karyotyping both use a cell culture test, only the origin of the sample is different, although both samples are obtained during an invasive procedure.

† The 'new gold standard' test regime was used because results from this study were to be used to determine the sensitivity and specificity of molecular tests. In addition, ethics considerations would have prevented use of molecular tests alone until sensitivity and specificity was defined.

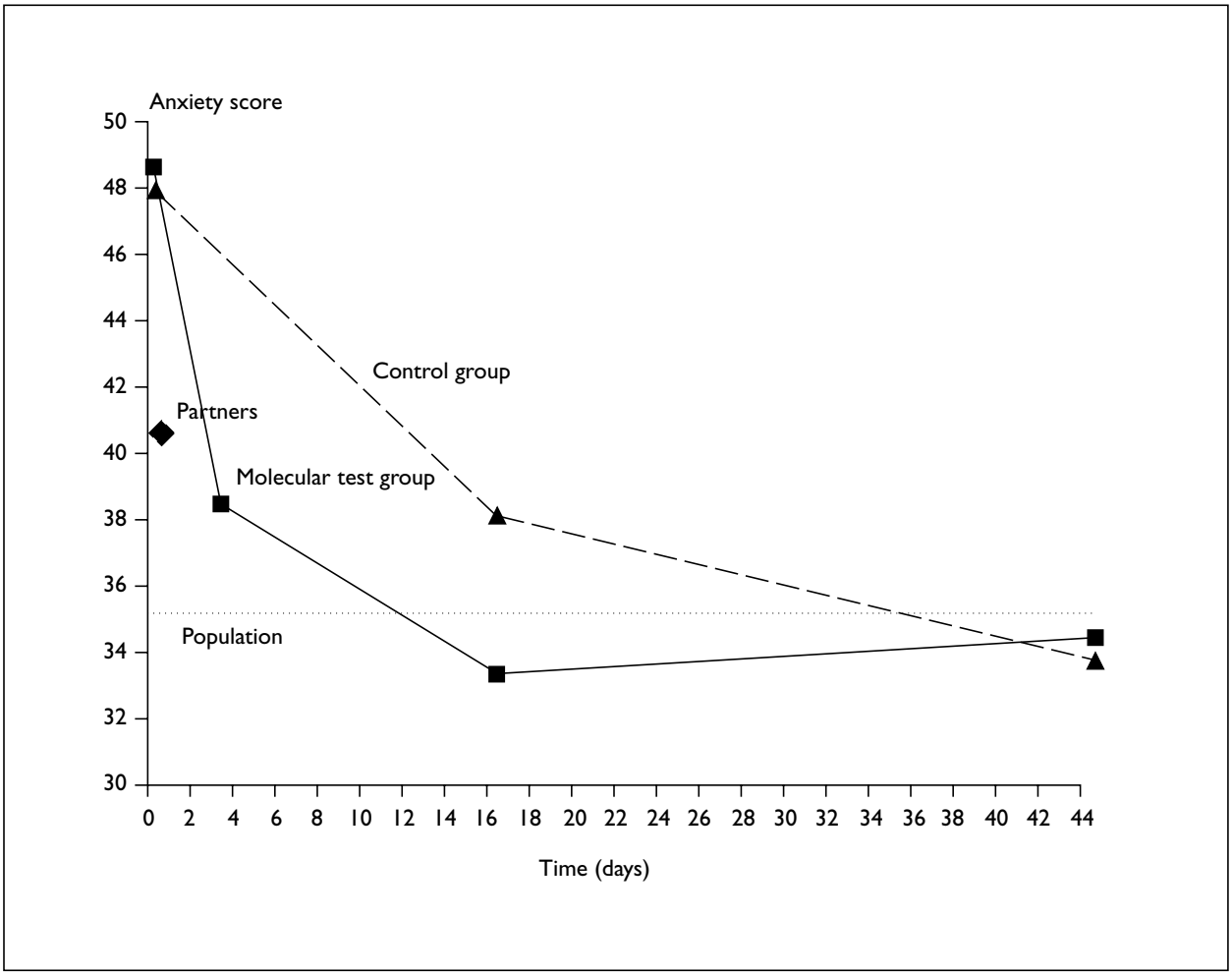


FIGURE 17 *Changes in anxiety scores throughout the testing process for molecular test and control groups*

(dashed line), the points were identical, except for the absence of (ii). The Anxiety State score for the general female population in the same age range (population score norm) is plotted as a horizontal line (score = 34.8).

The figure demonstrates that both intervention and control patients reported clinically significant anxiety at amniocentesis (scores greater than 42[69]). The difference in scores at amniocentesis between the control and intervention groups was not significant (Student's *t*-test, $p = 0.894$). On receipt of the first test result, however, whether it was the molecular test or karyotyping, a significant reduction in anxiety scores (Student's *t*-test, $p < 0.001$) was observed. This reduction was almost exactly the same for both groups, although the control group had to wait a mean of 13.5 days longer for their first test result (karyotyping). *Figure 17* also demonstrates that the intervention group (solid line) showed a further substantial drop in anxiety on receipt of the karyotyping result (Student's *t*-test, $p < 0.001$).

For the control group, anxiety continued to show a reduction ($p < 0.001$) at 4 weeks after receipt of their karyotyping test result. However, there was no corresponding significant change in anxiety for the intervention group between the karyotyping result and 4 weeks later (Student's *t*-test, $p < 0.001$). Therefore, equal reassurance appeared to be provided by receipt of the first result, either a karyotyping result (control group) or the molecular test result (intervention group).

For the molecular test group, women were reporting anxiety scores slightly below the population norm (not significant) by the time the second (karyotyping) test result was received (about 2 weeks after receipt of their first test result). By this time, the observed score was not significantly different from the population norm (Student's *t*-test, $p < 0.001$). For women undergoing the traditional karyotyping test, however, anxiety was not reduced to the population norm until 28 days (4 weeks) after receipt of their first (and only) test result.

For partners who agreed to complete a questionnaire (85/89), the mean anxiety reported at entry was 40.4. This was significantly lower than the women's initial mean score of 48.7 (Student's *t*-test, $p < 0.001$), but significantly higher than the population norm of 35.7 for the general male population in the same age range (Student's *t*-test, $p < 0.001$). Women and partners showed no trait difference from the general population at the $p = 0.01$ level (Student's *t*-test, $p = 0.986$ [men]; $p > 0.043$‡ [women]). Thus, an anxiety-specific outcome measure demonstrated significant differences between the intervention ('new gold standard') and control (karyotyping) groups, at specific points during the testing period and over several weeks beyond.

Parental health status

Changes in health status were also measured using a generic instrument (EuroQoL EQ-5D) over the testing period. The points at which measurements were taken were as for the Speilberger questionnaire. For women these points were over the whole testing process and beyond, and for partners only at the visit for booking or amniocentesis. Changes in women's health status are shown in *Figure 18* for the intervention group for the 138 (141 total responses) valid responses received; and in *Figure 19* for the 53 responses received in the control group.

Both figures illustrate the percentages of women reporting a reduction in any of the five dimensions measured (mobility, self-care, pain/discomfort, usual activities and anxiety/depression) at each point of measurement. For comparison purposes, values for the general population are also shown (first columns in *Figures 18* and *19*).

The figures show that self-care was not affected and the three dimensions that are affected during the testing period are:

- mobility and usual activities (with increasing numbers of women reporting some reduction on both these dimensions over time)

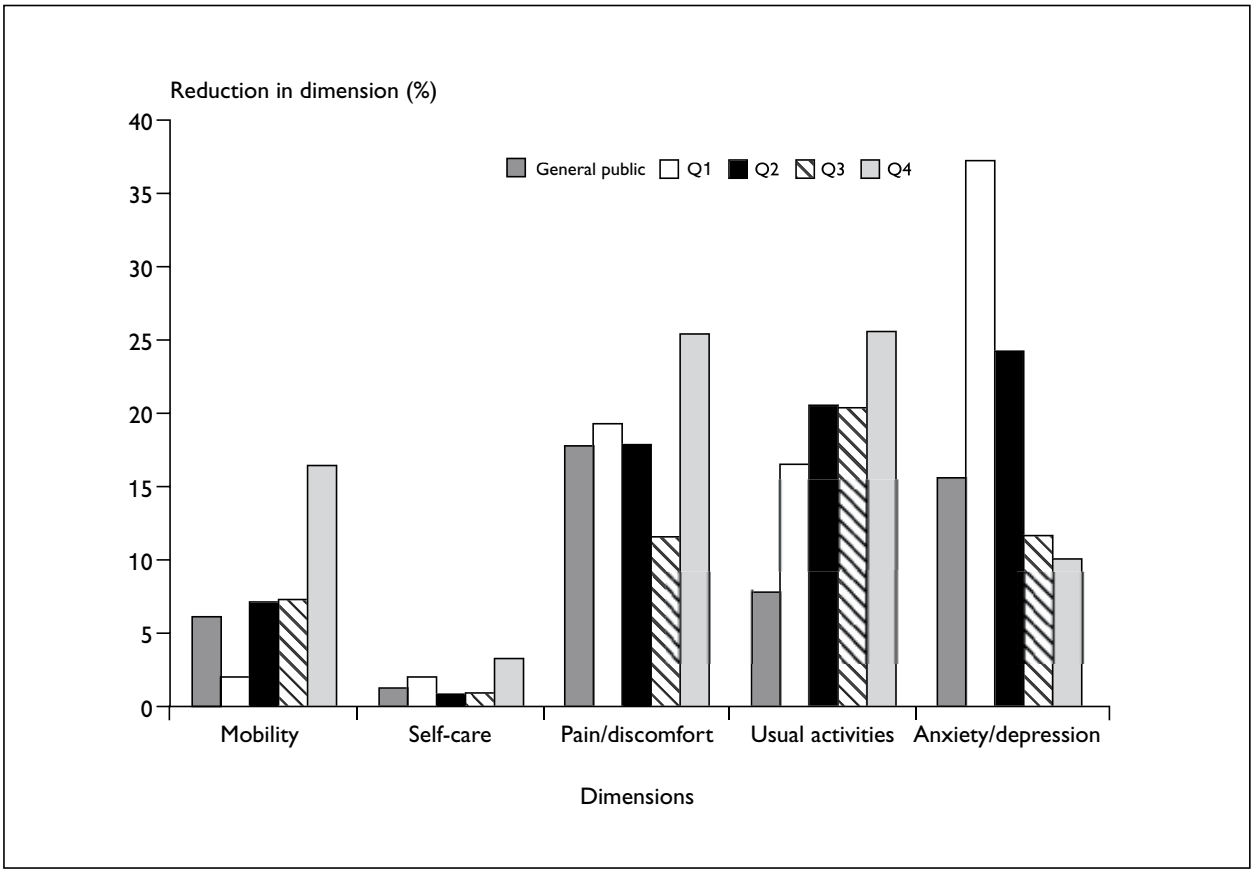


FIGURE 18 *The five dimensions of EuroQoL for the intervention group at different points during the testing process compared with those for the general public*

‡ Trait was measured at each point; $p = 0.043$ was the smallest *p*-value for the series.

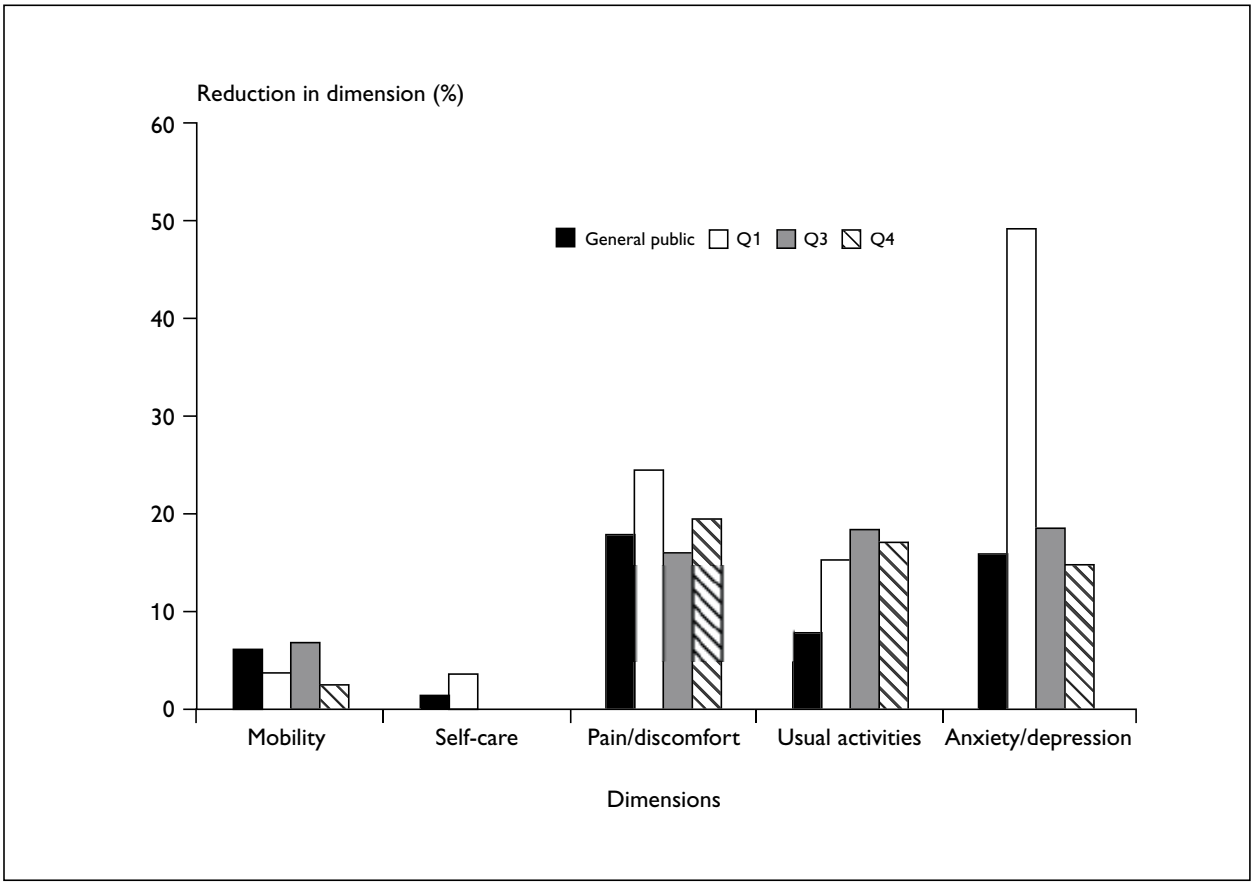


FIGURE 19 *The five dimensions of EuroQoL for the control group at different points during the testing process compared with those for the general public*

- anxiety/depression (with a decreasing number of women reporting some reduction in QoL on this dimension over the testing period)
- pain/discomfort (with an initial reduction in the number of women reporting some impact on this dimension, rising once again towards the end of the testing period).

P-values for mobility, usual activity and anxiety/depression are shown in *Table 44* for both control and intervention groups, based on the hypothesis that the study group was no different than the general public (paired *t*-test). At amniocentesis, and throughout most of the testing period, women's scores for mobility were not significantly different from those for the general public, although there was some indication that differences were beginning to emerge for the last questionnaire. This may be linked to advancing pregnancy, reflected in the fact that the last questionnaire was answered some 7–8 weeks after amniocentesis when women were between 21 and 26 weeks' gestation.

However, the scores for usual activities and anxiety/depression did differ significantly from those of the general public at the outset. In the

TABLE 44 *P-values for differences between health states for intervention and control groups compared to general public in same age range*

Time of test	Mobility		Usual activities		Anxiety	
	Intervention	Control	Intervention	Control	Intervention	Control
At amniocentesis	0.036	0.327	< 0.0001	< 0.0001	< 0.0001	< 0.0001
At receipt of molecular test result	0.550		< 0.0001		< 0.0001	
At receipt of karyotyping result	0.408	0.557	< 0.0001	0.011	0.170	0.388
4 weeks after karyotyping	< 0.0001	0.250	< 0.0001	0.025	0.097	0.525

former case (usual activities), this difference persisted throughout the testing period; in contrast, for anxiety/depression the score peaked for the first questionnaire, and then reduced to a value at the end of the testing process that was no different from that reported by the general public. Once again, it can be postulated that the difference in usual activity scores arises from advancing pregnancy rather than from the testing process itself. The pattern for anxiety/depression mirrors the state scores for the Speilberger Anxiety Inventory.

For the other two dimensions (pain/discomfort and self-care), there were no significant differences between the two study groups and the general public across the testing period.

Direct comparison of the intervention and control groups, however, demonstrated that observed differences on individual dimensions were not statistically significant at the points at which they could be compared – prior to amniocentesis, at receipt of first result and 4 weeks after the karyotyping result. Furthermore, the only individual dimensions on which significant differences could be detected between questionnaires were on the dimension of anxiety/depression and the last mobility score.

Also incorporated in the EuroQoL questionnaire is a 'thermometer' visual analogue scale (VAS) allowing women to give their own self-rated health status. Mean self-rated health status for females aged 30 to 39 is 85.9 for the UK population. The results for both the intervention group and the control group are shown in *Table 45* and can be compared to this value.

Women in both the molecular test group and the control group reported significantly worse overall health status at amniocentesis (mean scores of 80.3 and 80.8, respectively, Wilcoxon signed rank test, $p < 0.001$), but a better than UK population mean health status upon receipt of results (scores 87.0 and 87.6, respectively). Partners' mean health status (only recorded at the time of amniocentesis or booking the test) was found to be 84.7 for the control group and 86.0 for the intervention group. These values were not significantly different from those of women in each respective group.

The recorded change in self-rated health status between amniocentesis and receiving the first test result was 3.13 for the intervention group ('new gold standard') and 3.39 for the controls (karyotyping). All these values are illustrated graphically in *Figure 20*. Health status was not significantly different in the intervention group between those with different test preferences, or (in intervention and control groups) other dependent variables such as age, ethnicity or parity.

The Wilcoxon signed rank test was used to compare the differences (two-tailed) between the two groups in self-rated health status at each test point. This analysis was conducted for those cases with a complete set of responses: 90 women in the intervention group and 40 cases in the control group. There were no significant differences in age, parity or ethnic origin between women with a complete set of responses and the remainder. The *p*-values (χ^2) relating to changes in health status, recorded at different points in the testing period (as illustrated in *Figure 20*) are given in *Table 46* for both the intervention and control groups.

Significant improvements in health status were recorded for the intervention group across the main testing period, compared to their health status measured at the outset (i.e. amniocentesis). Improvement in health status between receipt of the molecular test result and karyotyping was also significant (χ^2, $p = 0.003$), but there was no significant improvement between the receipt of karyotyping test result and 4 weeks after.

TABLE 45 *Women's changing self-rated health status for molecular test group and traditional test group during testing process*

Time (days)	Number in molecular test group ('new gold standard')	Health status mean (quartiles)	Number in control group	Health status mean (quartiles)
0	138	80.3 (75.0–92.0)	53	80.8 (72.5–93.5)
3	111	83.4 (78.0–94.3)		
16.8	93	86.8 (80.0–96.5)	43	84.2 (70.0–98.0)
44.8	90	87.0 (80.0–95.0)	40	87.6 (81.2–95.0)

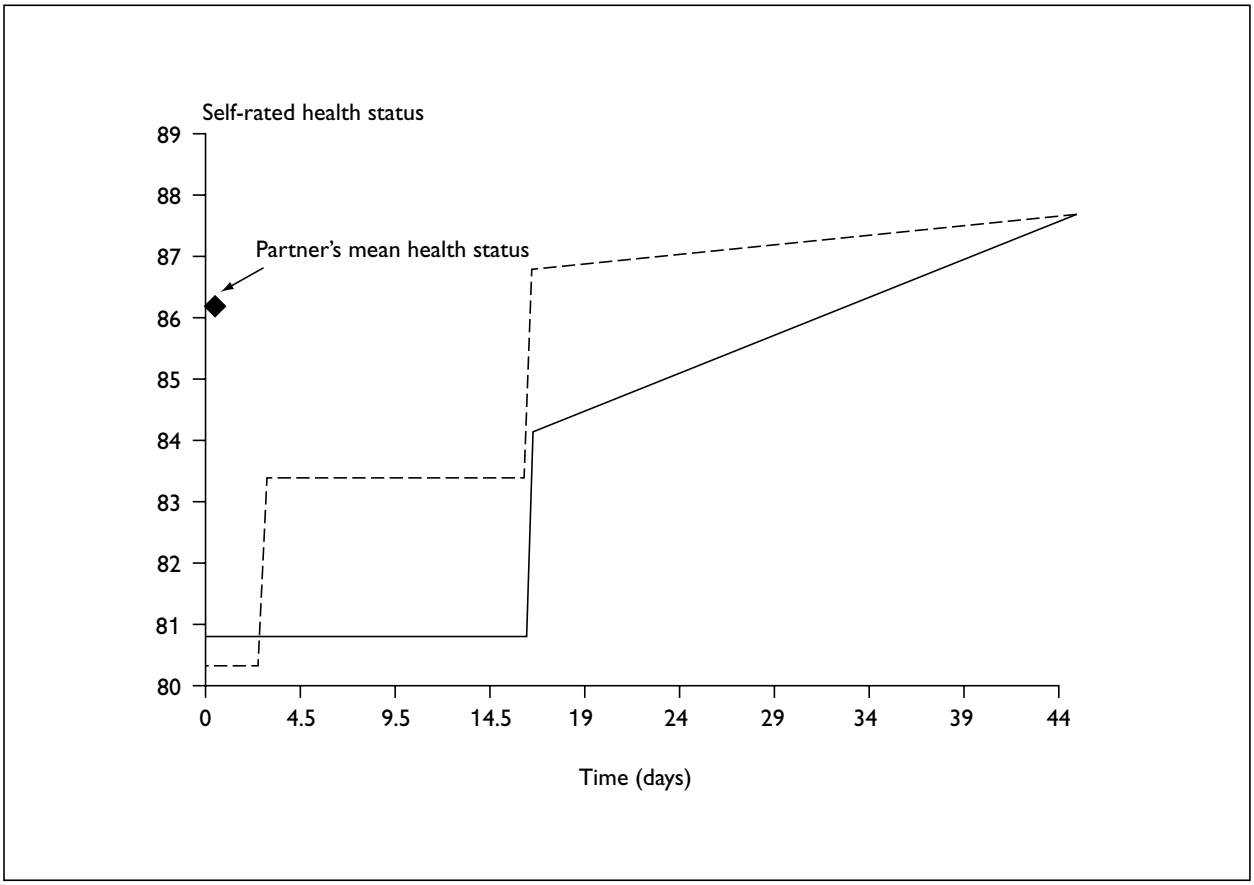


FIGURE 20 *Self-rated mean health status scores (VAS) for women in the intervention and control groups and (point measurement) for partners (– – –, intervention group; ——, control group)*

TABLE 46 *P-values (χ^2) comparing differences in self-reported health states for various intervals for both control and intervention groups*

Time of test (days)	Intervention group (90 cases)			Control group (40 cases)	
	Test day (0)	Molecular result (3)	At karyotyping result (~ 17)	Test day (0)	At karyotyping result (~ 17)
At receipt of molecular result	0.001				
At receipt of karyotyping result	< 0.001	0.003		0.522	
4 weeks after karyotyping result	< 0.001	0.057	0.155	0.019	0.97

This suggests that most of the benefit (i.e. reassurance) resulted from the receipt of the molecular test result, as might be expected. A similar pattern was recorded for the control group, although the only significant difference was between receipt of karyotyping result and 4 weeks later ($p = 0.019$).

Cost–utility analysis

It is possible to calculate cost–utility ratios using the health status differences reported above and the costs recorded for the two testing regimes. Using the cost information in *Table 14* and the mean health status scores in *Figure 20,* it is possible to estimate a marginal cost per quality-adjusted life-year (QALY) gained for the 'new gold standard' test regime (using a 5-probe or multiplex molecular test) compared to the status quo (karyotyping). Furthermore, if an assumption is made that similar changes in health status would be observed for a 'new gold standard' test regime that used a Down only molecular test, then the marginal cost per QALY gained can also be estimated for this form of 'new gold standard' testing regime.

In general, public valuation of EuroQoL health states is based on asking individuals to consider that the subject remains in a particular health state for a period of 1 year. In the present case, changes in health status were measured over a period of 45 days. Therefore, the health status measure used in the cost–utility calculation was the self-rated health status score entered on the VAS included in the questionnaire. The health status gain for those receiving the molecular test, expressed in QALYs and taking a full year as 365.25 days, is calculated to be 0.00179 QALYs over a period of 45 days.

Estimates of the marginal cost per QALY gained for the four 'new gold standard' test options are given in *Table 47*. Results indicate that use of multiplex Q-PCR tests and both types of Down only molecular test produce similar cost–utility ratios, at between £23,500 and £26,500 per QALY gained. It is calculated that use of 5-probe FISH tests would result in an even higher ratio (£41,940 per QALY gained).

The first priority in evaluating molecular tests was to measure sensitivity and specificity of the tests in the service setting. This could only be assessed ethically by performing molecular tests prospectively, together with karyotyping. Thus, the prospective trial carried out was only able to measure changes in health status associated with the use of molecular tests as an add-on to karyotyping (i.e. women received the 'new gold standard' test). Further studies will be required to measure the changes in health status associated with molecular tests used as a replacement for karyotyping, or for other testing regimes in which molecular tests partially replace karyotyping. In order to assess the cost-effectiveness of these latter testing regimes, the pragmatic approach taken in chapter 7 and in the remainder of this chapter was adopted.

Extending the cost-effectiveness analysis

Simple cost-effectiveness ratios (cost per case detected) were presented in chapter 7 for a range of test regimes in which molecular tests partially or fully replace karyotyping. These were based on cost per case detected.

A more comprehensive way of assessing the cost-effectiveness of these different testing regimes would be to weight the cost-effectiveness ratios to account both for the length of time women have to wait for results and for the information content of the test result once received. This would acknowledge the fact that it is not sufficient to only consider how many 'cases' a test detects, but it is also important to consider when these cases are identified. Thus, if a test for Down syndrome only is conducted, uncertainty about other abnormalities will remain until the end of the pregnancy. Conversely, a testing regime that provides women with a molecular test for the five most common chromosome abnormalities in addition to karyotyping ('all common' 'new gold standard') will remove most uncertainty at the time of the molecular test result, and remaining uncertainty about rarer abnormalities will be removed with the karyotyping result after approximately 3 weeks. The latter testing regime will inevitably incur a higher cost but it will detect more abnormalities prenatally from the same population and provide information more rapidly. Ideally, all three factors should be reflected in any cost-effectiveness calculation, although the simple cost-effectiveness ratios presented in chapter 7 only allowed for the first two of these factors. Before these ratios can be weighted to make allowance for the degree of uncertainty remaining after a test result, and the time spent in different states of uncertainty, it is necessary to first calculate the turnaround time for different types of test result.

Time taken for production and receipt of molecular test results

The mean time taken from reception of specimens at the laboratory to reporting of a negative result is given in *Table 48* for molecular tests and karyotyping. (Those with a positive result after the molecular test left the study group at this point. For these patients karyotyping was reported as soon as cell cultures were available.) As can be seen from this table, for the intervention group

TABLE 47 *Cost per QALY for test regimes using molecular tests*

Test regime		Molecular test	Marginal cost per QALY gained
'New gold standard'	Down only	FISH	£23,542
		Q-PCR	£24,670
	All common abnormalities	FISH	£41,940
		Q-PCR	£26,514

TABLE 48 *Mean time and range, in days, from reception at laboratory to posting result to patient*

Type of test	Mean days in laboratory (range)	
	Northern Ireland	West Midlands
Karyotyping	20.8 (13–27)	12.8 (7–17)
Molecular tests (FISH: Northern Ireland; Q-PCR: West Midlands)	1.1 (1–2)	4.4 (2–8)
Difference between two test types	19.7	8.4

of 67 patients in Northern Ireland and 74 in the West Midlands, use of molecular tests considerably reduced the laboratory time required for processing samples.

However, the turnaround times recorded in the Northern Ireland and West Midlands laboratories differed. The explanation for this may be that, for molecular tests, the time required for processing samples in a service setting will be sensitive to the day on which samples are received in the laboratory. It was observed during the study that, for the series of cases in Northern Ireland, all samples were received from one clinic early in the working week. This was not the case for the West Midlands samples, which were derived from clinics spread over the whole working week, reflecting the different working practices of different obstetricians. Thus, the value of 1.1 days in Northern Ireland represents a best case scenario, where a laboratory receives samples from all its antenatal clinics early in the working week and can therefore complete processing during that same week. In contrast the observed value of 4.4 days in the West Midlands represents a worst case scenario. In this case the laboratory received trial samples from antenatal clinics spread across the week and these samples were only processed during Monday to Friday because the trial staff did not start processing during the weekends. From the survey of UK laboratories, 57% of genetics laboratories work on a Saturday and a third are open for a period on a Sunday. In practice, therefore, the turnaround time for molecular tests in UK laboratories is likely to be somewhere between the two extremes shown in *Table 48*.

Even if laboratory processing of molecular tests from receipt to result only takes 1–2 days in service conditions, certain routes for transferring results to women could potentially substantially increase the overall time proportionately that women have to wait for a test result, particularly where results are reported via clinicians.

For karyotyping, the total test wait time (i.e. between amniocentesis and a patient receiving their test result) reported in the UK laboratory survey varied between sites. Where results were reported via antenatal clinic staff, additional test wait time appeared to be introduced. This reporting practice was common in the vast majority of Trusts for negative results; 26/30 of the laboratories surveyed posted the results of negative tests to the referring clinician who then reported to the patient. In only one centre (West Midlands) were negative results posted directly to women by the laboratory. In a further three centres (including Northern Ireland), it was established practice for the patient to contact the cytogenetics laboratory by phone for results. Although the latter practice might minimise the delay in transmitting results, it is difficult to envisage this being feasible for a very large laboratory.

No respondents in the UK survey indicated that results were transmitted electronically to the referring clinician, although electronic links already exist for other diagnostic services such as biochemistry and pathology, with results being transmitted via NHSNet or other dedicated systems. Positive results were always reported via antenatal clinic staff or a member of the genetics services.

Estimating weighted cost-effectiveness ratios

As discussed on page 19, 'Sensitivity and specificity of molecular tests', molecular tests (particularly those configured to detect Down syndrome only) inevitably will have a lower relative sensitivity and specificity than karyotyping. This means that if a molecular test were to be used on its own, some level of uncertainty would remain until the baby is born, compared to karyotyping. If it is assumed that once a baby is born the presence and nature of any abnormality will be confirmed,[§] then there will be no remaining uncertainty about a diagnosis.

[§] Many of the less significant abnormalities may not be evident at birth or, in some cases, even later in life.

TABLE 49 *Number of uncertainty days for 'new gold standard' test regime*

Test regimes		Molecular test	Direct cost per patient tested	Uncertainty	Days in state	Uncertainty days
'New gold standard'	Down only	FISH	£131.82	1 0.3522 0.0001	2 15 158	7.30
		Q-PCR	£133.84	1 0.3804 0.0001	2 15 158	7.72
	All common[a]	FISH	£164.75	1 0.1395 0.0001	2 15 158	4.11
		Q-PCR	£137.14	1 0.1766 0.0001	2 15 158	4.66
Karyotyping (status quo)			£89.68	1 0.0001	17 158	17.02

[a] *Trisomies 21, 18 and 13 and X,Y chromosome abnormalities*

For the intervening period, however, a parameter best described as **uncertainty days** can be estimated for each of the test configurations. This can be calculated by combining two values: (i) relative sensitivity and specificity (*Table 12*), to express the degree of uncertainty about the presence or absence of chromosome abnormalities, and (ii) days spent in each state of uncertainty (*Table 49*).

Uncertainty days, calculated in this way (compared to the traditional test), are presented in *Table 49* for the molecular tests as add-on (i.e. 'new gold standard' test regime), and in *Table 50* for the molecular tests as replacement for karyotyping (i.e. 'molecular test for all' test regime).

Appendix 6 provides information on the detailed calculations performed.

As can be seen from *Table 49,* use of karyotyping is estimated to result in the largest number of uncertainty days, when compared to the various 'new gold standard' test regimes. This is because, if a 'new gold standard' test regime is used, almost all the information provided by karyotyping is available to parents very early in the testing process.

In contrast, *Table 50* shows that for the 'molecular test for all' regimes (i.e. molecular tests as a replacement for karyotyping), the reduced information provided by Down only

TABLE 50 *Number of uncertainty days for 'molecular test for all' regime*

Test regimes		Molecular test	Direct cost per patient tested	Relative sensitivity	Uncertainty	Days in state	Uncertainty days
Molecular for all replacing karyotyping	Down only	FISH	£42.14	0.6478	1 0.3522	2 173	62.93
		Q-PCR	£44.16	0.6196	1 0.3804	2 173	67.81
	All common[a]	FISH	£75.07	0.8605	1 0.1395	2 173	26.13
		Q-PCR	£47.46	0.8234	1 0.1766	2 173	32.55
Karyotyping (status quo)			£89.68	0.99	1 0.0001	17 158	17.02

[a] *Trisomies 21, 18,13 and X, Y chromosome abnormalities*

molecular test results in the maximum number of uncertainty days for these tests. However, 5-probe FISH and, to a slightly lesser extent, multiplex Q-PCR (both configured for the five common abnormalities) approach the value for karyotyping. In these last two cases, any loss of information is compensated for by a much speedier result.

It should be noted that the values in *Table 50* are estimated from the clinician's perspective; uncertainty is calculated against perfect knowledge of how each test performs. In other words the clinician will be aware that even karyotyping can give a false result in some cases, as shown by the relative sensitivity of less than unity in *Table 50* (see also page 19, 'The gold standard, karyotyping').

If the simple cost-effectiveness ratios (cost per case detected) reported in chapter 7 are weighted by the uncertainty days in *Tables 49* and *50*, then a modified cost-effectiveness ratio can be estimated as given in *Table 51*. Once again, 'cases detected' are defined as all chromosome abnormalities detectable by the test configuration used and the figures given for 'molecular test for all' assumes that karyotyping will be replaced by molecular tests.

Table 51 presents weighted cost-effectiveness ratios for the eight main test configurations. These weighted ratios allow for the relative costs of the test, the numbers of cases detected, and the relative time in which women remain in different states of uncertainty. Based on these ratios the 'new gold standard' test regime option would now appear to represent better value than karyotyping on its own. This is because the initial (partial) information on abnormalities available quickly through the molecular tests plus the remaining information provided by karyotyping combine to give significantly better value than karyotyping alone. For the 'new gold standard' test regime it appears that the 'all common' molecular tests are the best tests to use, with their higher cost offset by the increased number of cases detected early. The extra information provided by the 'all common' form of the 'new gold standard' is reflected in the reduction of around 56–58% in the cost-effectiveness ratio compared to karyotyping.

However, if Down only molecular tests are used for the 'new gold standard', it appears that the loss of early information is not fully compensated for by the reduced cost, and the weighted cost-effectiveness ratio is calculated to be less favourable – a reduction of 32–37% compared to karyotyping. The benefit of more rapid reporting is therefore undermined by the loss of information from the Down only test configuration, with fewer cases detected early, and a higher degree of uncertainty remaining until the karyotyping result.

Unlike add-on tests, the use of molecular tests as a replacement for karyotyping ('molecular test for all' regime) is calculated to be a less cost-effective strategy than karyotyping, substantially increasing the weighted cost per case detected (by between about 55% and 412%). This is because limited information is available with all molecular tests and therefore high levels of uncertainty about other abnormalities remain to the end of pregnancy. Even with the extra information provided by the 'all common' form of the 'molecular test for all'

TABLE 51 *Comparative cost-effectiveness of tests weighted for uncertainty days*

Test configuration			Ratio of uncertainty days to karyotyping	Cost per case detected (% change from status quo)	
				Unweighted	Weighted
'New gold standard'	Down only	FISH	0.36	£3,184 (+47%)	£1,366 (–37%)
		Q-PCR	0.38	£3,232 (+49%)	£1,467 (–32%)
	All common[a]	FISH	0.20	£3,979 (+84%)	£960 (–56%)
		Q-PCR	0.23	£3,312 (+53%)	£908 (–58%)
Karyotyping (status quo)			1.00	£2,166 (0%)	£2,166 (0%)
'Molecular test for all'	Down only	FISH	2.66	£2,657 (+23%)	£9,826 (+354%)
		Q-PCR	2.87	£2,784 (+29%)	£11,096 (+412%)
	All common[a]	FISH	1.11	£2,768 (+28%)	£4,250 (+96%)
		Q-PCR	1.38	£1,750 (–19%)	£3,347 (+55%)

[a] *Trisomies 21, 18, 13 and X, Y chromosome abnormalities*

regime there is a calculated increase of between 55% (Q-PCR) and 96% (FISH) in the cost-effectiveness ratio compared with karyotyping.

As explained, the values in *Table 51* are based on the assumption of 'perfect knowledge' of the test and its performance (or the clinician's perspective). These values may therefore change if calculated from a patient's perspective, dependent on what information individuals expect from the test result. In particular, if women expect, and want, only a test result for Down syndrome, the relative values would change as shown in *Table 52*.

Table 52 shows values for the weighted cost-effectiveness ratios based on the assumption that a test for Down syndrome alone is requested or expected. As can be seen, in this situation the 'molecular test for all' regimes are more cost-effective than karyotyping or the 'new gold standard'. The Down only test configuration of the 'molecular test for all' regime is the most cost-effective. This is because all molecular tests are technically equivalent in performance for the detection of Down syndrome and the 'new gold standard' and the 'all common' configuration of the 'molecular test for all' regime have a higher cost per sample tested than the Down only 'molecular test for all' regime.

In the situation depicted in *Table 52,* the unweighted cost per case detected also changes compared to *Table 51,* because the intention is only to detect Down syndrome and thus the number of abnormalities to be detected falls by 60%. Therefore, for karyotyping, the unweighted cost per case detected rises by 161%, from £2166 to £5654 per Down syndrome case detected. This reflects the relative inefficiency of this test, which is, in effect, over-diagnosing. Similarly, the unweighted cost per case of Down syndrome detected in *Table 52* for the 'molecular test for all' regime rises for the 'all common' test configuration, while remaining the same for the Down only configuration compared to *Table 51.*

Table 52 also shows that in a situation where women are only expecting a test for Down syndrome, use of molecular tests alone is seen to be much more cost-effective than use of the 'new gold standard'. This reflects the more expensive test cost of the 'new gold standard' regime compared to the information expected. Although all molecular test configurations demonstrate better weighted cost-effectiveness ratios than karyotyping, because the test results are received within 2 days rather than 13–21 days, the Down only 'molecular test for all' regime (FISH or Q-PCR) is the most cost-effective as it minimises waiting and has the smallest direct cost (weighted cost per case detected, 95% lower than for karyotyping).

The assumptions on which the weighted cost-effectiveness calculation are based are untested and need validation. In their current form, however, the assumptions highlight the sensitivity of the cost-effectiveness approach in terms of the outcomes chosen. In the unweighted cost-effectiveness it is assumed that the outcome of the number of cases detected is the sole outcome. Even this simple outcome is not without problems as it may be asked what outcomes the women tested were expecting and whether there is a complete match in expectations between clinician

TABLE 52 *Comparative cost-effectiveness of tests weighted for uncertainty days (for Down syndrome only)*

Test configuration			Ratio of uncertainty days to karyotyping	Cost per case detected (% change from status quo)	
				Unweighted	Weighted
'New gold standard'	Down only	FISH	0.10	£8,311 (+47%)	£902 (–84%)
		Q-PCR	0.10	£8,439 (+49%)	£916 (–84%)
	All common[a]	FISH	0.10	£10,388 (+84%)	£1,128 (–80%)
		Q-PCR	0.10	£8,647 (+53%)	£939 (–83%)
Karyotyping (status quo)			1.00	£5,654[b] (0%)	£5,654 (0%)
'Molecular test for all'	Down only	FISH	0.10	£2,657 (–53%)	£288 (–95%)
		Q-PCR	0.10	£2,784 (–51%)	£302 (–95%)
	All common[a]	FISH	0.10	£4,733 (–16%)	£514 (–91%)
		Q-PCR	0.10	£2,992 (–47%)	£325 (–94%)

[a] *Trisomies 21, 18, 13 and X, Y chromosome abnormalities*
[b] *Cost per case of Down syndrome detected*

and patient, where for example the women has entered a screening programme for Down syndrome. The weighted cost-effectiveness analysis modifies some of the conclusions, but at the expense of invalidated extended outcome measures.

Valuing preferences: the cost–benefit approach

Cost-effectiveness analysis can only provide information about the relative value of the tests in terms of their technical efficiency, while cost–utility analysis only measures the impact tests have on health status. Cost-effectiveness has already been explored in chapter 7 and the impact of molecular tests on health status is explored on page 81, where *Table 47* presents a marginal cost–utility analysis.

Neither of the above analyses can provide an insight into how patients or clinicians value different tests, what their test preferences are, or how strongly they feel about their preferences. An alternative approach, which can do this, is to explore costs and benefits within a Paretian framework using cost–benefit analysis. This form of economic analysis was developed within transport economics as a means of both valuing alternatives and collecting evidence on strengths of preferences, and it has been used widely in healthcare evaluations of diagnostic interventions.[70–73] The cost–benefit analysis technique most commonly used in medical diagnostics, estimating WTP, involves measuring how much a respondent is prepared to pay from their own income, or how much income the respondent would be willing to forego, to receive the diagnostic test of their choice. The basis for this approach is that an individual will value a service or item in the context of their general spending and thus will offer insight into the importance of the service or item by the amount they are prepared to pay. In the present study, the value of the new molecular tests to various stakeholders was explored using such an approach.

Women, their partners, health commissioners and the general public were all surveyed to explore their valuation, in terms of both preferences and WTP, for the different test alternatives. A similar instrument, based on contingent valuation,[¶] was developed and piloted for each population. For women, the only opportunity for collecting information on *ex ante* preferences was prior to amniocentesis. As a condition of the Multicentre Research Ethics Committee approval the patient had to be consented to the trial and any data collection instrument administered by the counselling midwife, or obstetrician, during their pre-amniocentesis counselling session. This occurred either at the booking visit for amniocentesis or immediately before amniocentesis, depending on source of referral and local practice. Therefore, a simple questionnaire for eliciting WTP, consistent with reliability and validity and within the contingent valuation framework, was required for use in the study. A 1993 study provides a useful analysis of the relative merits of open-ended or bidding and closed-ended bidding systems.[74] In particular, for self-administered questionnaires there are advantages to properly piloted closed-ended questions and this model was therefore chosen for the present study.

General public WTP

Following ethical approval, 1000 people aged 16 and over were randomly selected from a Health Authority register (Solihull) for survey (see page 58, 'General public'). The WTP question was first piloted on 12 staff within the University of Warwick. After initial amendments a further draft was sent out to 55 members of the public within Warwickshire.

Responses to the survey, conducted during May 1999, were received from 289 people (28.9% response rate) after one reminder letter was sent out; 41% of respondents were male and the age and income ranges are shown in *Figures 21a* and *21b*. Respondents predominantly came from an older population and from higher income brackets and 16% had no children.

The amount respondents were willing to pay is presented in *Figure 22*. The data is ungrouped in *Figure 22a* and grouped in *Figure 22b*. The distribution of values that respondents were willing to pay was bimodal, the mean value being £113 (median £55). Because the distribution of values of WTP was not normal for any of the surveys, median values were used for comparison throughout.

As reported on page 58 ('General public'), a majority of the general public (59%) chose

[¶] For pragmatic reasons contingent valuation was used rather than conjoint analysis, although the latter has been used in other studies exploring antenatal screening in the UK.[71] The basis for using contingent valuation and discussion of the use of cost–benefit analysis in this setting is included in appendix 7.

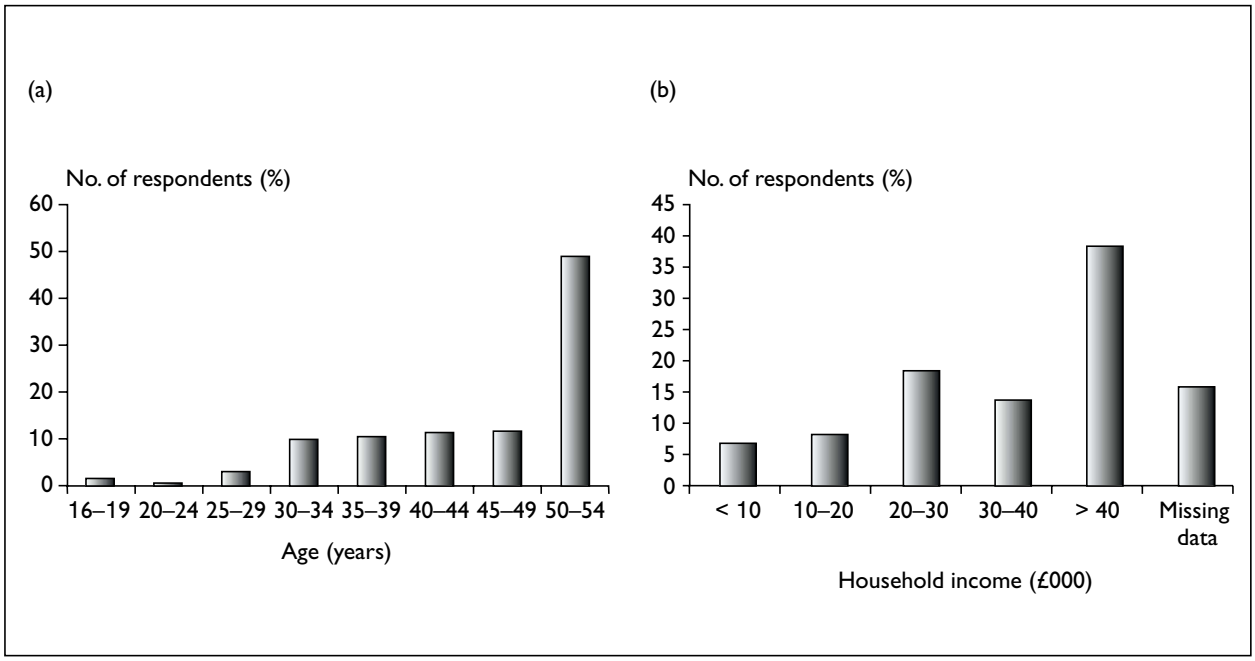


FIGURE 21 *(a) Age ranges and (b) household income of respondents in survey of general public*

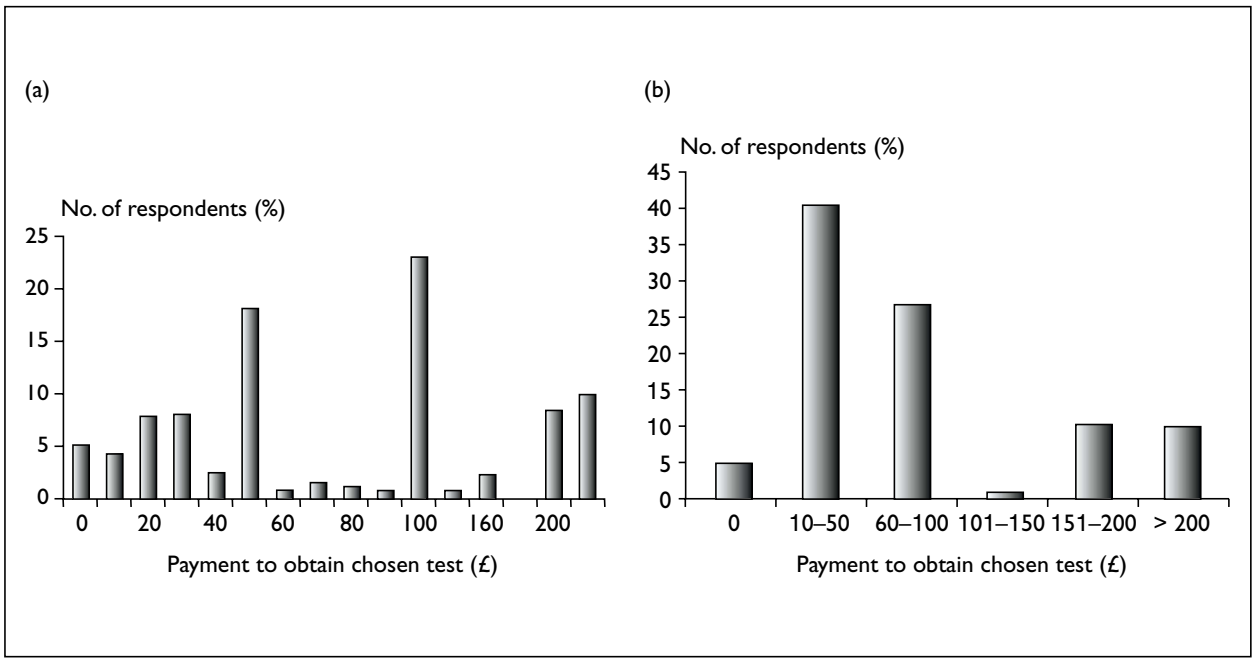


FIGURE 22 *General public WTP for their chosen test, (a) ungrouped and (b) grouped*

karyotyping in preference to a molecular test. Those choosing karyotyping reported a median WTP value of £100 and those choosing molecular tests a median of £50. The difference in median values is significant at the $p = 0.05$ level (Mann–Whitney, $p = 0.033$), indicating that those choosing karyotyping had a stronger preference than those choosing molecular tests. Test choice and WTP values were not sensitive to age, income or gender between test choice groups. It should also be noted that 0.7% of those surveyed did not respond to the WTP question and 2% of respondents indicated they would not accept a test at all (total 2.7%).

Parents WTP

The sample of women consenting to the prospective study has already been described on page 46 ('Recruitment of patients'). Of these individuals, only those offered the molecular test (the intervention group), and their partners, were

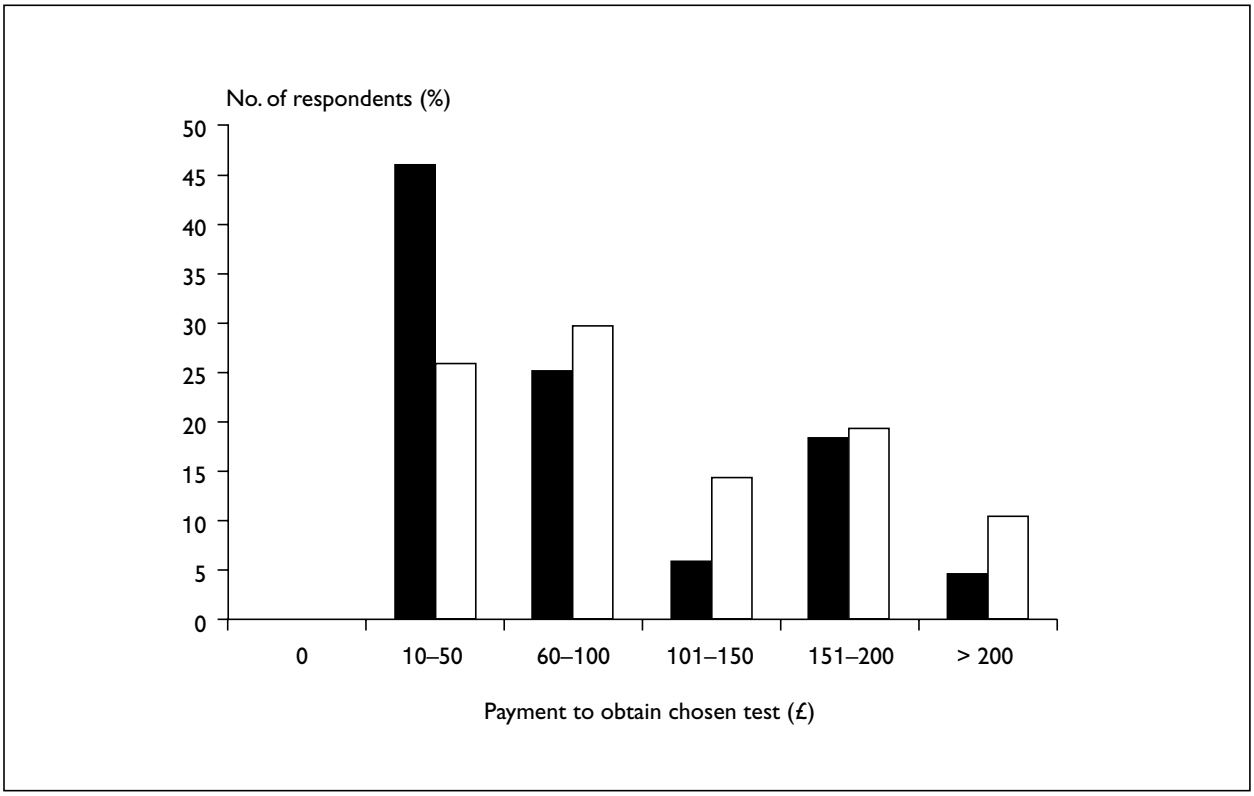


FIGURE 23 *Women's WTP for their chosen test, grouped (■, amount willing to pay; □, most willing to pay)*

asked about test preference and WTP for that preference in the study questionnaires.** The question used was piloted before use with clinical geneticists, midwives and obstetricians. The Regional Obstetricians and Gynaecologists Group were also asked for comment. Response rates for the questionnaires are given in *Table 25,* and women's characteristics in *Table 24*.

The distribution of amounts women reported they would be willing to pay to obtain their chosen test is shown in *Figure 23,* grouped as for *Figure 22b*. In addition, women were asked for the most they would be willing to pay and this is also illustrated; 38% of women did not answer the question on WTP and 45% failed to indicate the most they would pay.

As reported on page 57 ('Parents'), at the time of booking for amniocentesis, or the visit for amniocentesis, the majority of women stated a preference for molecular tests (66.7%). Although all women answered the test preference question, only 87 respondents (61.7%) answered the WTP question. Anecdotally, the midwives consenting women and their partners to the trial reported that several partners thought that there was a hidden agenda to this question and that the health service might be conducting market research to ascertain whether the test could become a charge to patients. It also appeared to be women accompanied by their partners who failed to answer the WTP question.

The median value women were willing to pay *ex ante* for their stated preference was £90 and 70% of women reported they would not be prepared to pay more than £100 to obtain their chosen test. *Ex ante*, those who preferred the molecular test were willing to pay a median value of £70 to obtain their choice. This value was not significantly different (Mann–Whitney, $p = 0.324$) to that for those who would prefer karyotyping, median £95. The distribution of WTP values for women undergoing the test was unimodal, unlike the distribution of values for the general public.

As well as asking women for their valuation of their test choice before undergoing amniocentesis,

** The question relating to WTP was as follows: "So that we can judge how strongly you feel about your choice, in question 5 we want you to imagine you have to pay for the test you have chosen. How much would you be prepared to pay to get your choice?".

women were also asked 4 weeks after their final result. At this point, fewer women (51% vs 67%) stated a preference for molecular tests (McNemar test, $p = 0.019$). The median WTP value at the end of the testing period was slightly increased (non-significant) to £100 for those choosing karyotyping. For those choosing the molecular test, although the median value had risen slightly, from £70 to £100, indicating an increasing strength of preference for molecular tests rather than karyotyping, this change was not significant (Mann–Whitney, $p = 0.336$). The majority of women who responded to both the first and the last questionnaire, 65% of respondents, did not change their minds on their WTP value (66%) between their *ex ante* and *ex post* responses, and differences in values for those who did were not significant (Wilcoxon signed rank test, $p = 0.522$). Furthermore the amount, in terms of WTP, was not related to age or parity.

In terms of test preference, there was a non-significant difference between women and their partners, with the partners also favouring the molecular test, although they were slightly less likely to choose these tests *ex ante* (51% vs 67% for women). *Figure 24* shows the amount partners reported they were willing to pay at amniocentesis, although only 42.7% of those partners who returned a questionnaire responded to the question on WTP. The median WTP value for partners of £60 was significantly less than that for women (Mann–Whitney, $p = 0.007$). Partners were willing to pay a median value of £125 for the choice of karyotyping compared to £50 for the choice of a molecular test, but this difference was not significant (Mann–Whitney, $p = 0.180$). Nevertheless, the difference implies that those who choose karyotyping had a stronger preference for their choice than those who chose the molecular test. The ratio of WTP values stated by fathers (2.5:1), indicating more than double the strength of preference for karyotyping, was even higher than that reported by the general public (2:1). This contrasts with women, who reported a 1.4:1 ratio *ex ante* and a 1:1 ratio *ex post*. All WTP values are summarised in *Table 53*.

Open text responses from partners in the questionnaire more closely matched those of the general public, who similarly valued the traditional test more highly and, according to the free text on the questionnaires, reported that they placed a high value on the additional information it provided. Furthermore, those doing the counselling for these couples anecdotally reported that partners placed a higher value on the information obtained from the test.

Health commissioners WTP

As part of the study, all health commissioners in the UK were surveyed (see page 58, 'Health commissioners'). In the survey questionnaire they

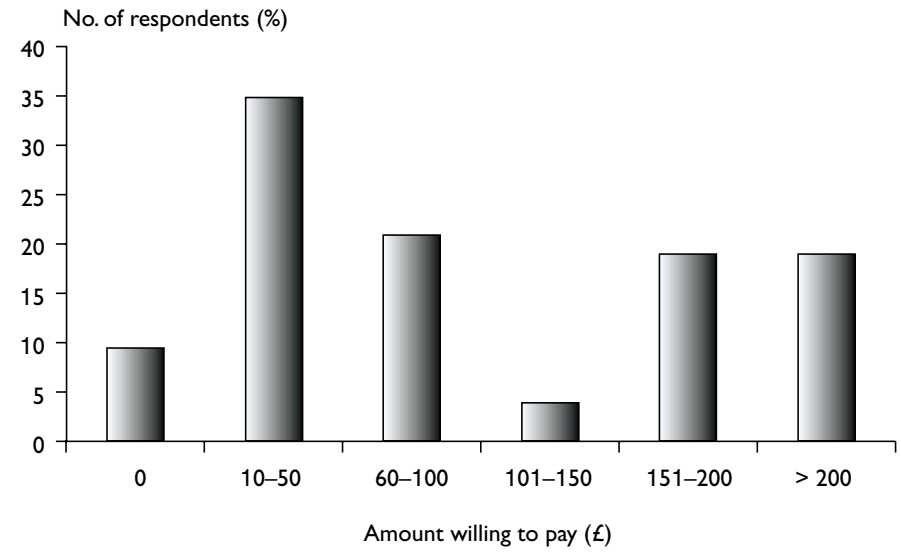


FIGURE 24 *Partners WTP for their chosen test, grouped*

TABLE 53 *Test preferences and WTP for preference for all groups surveyed*

Survey	Molecular tests		Karyotyping		*P*-values for difference		
	% with preference	Median WTP, £	% with preference	Mean WTP, £	Preference	Mean WTP, £	% with other preferences
Mothers *ex ante*	66.7	70	31.9	95	< 0.001	0.324	1.4
Mothers *ex post*	50.5	100	41.8	100	0.02	NS	7.7
Partners	50.8	50	42.6	125	NS	0.180	6.6
General public	37.8	50	59.4	100	< 0.001	0.033	2.8
Health commissioners	28.9	100	28.9	100	NS	NS	42.2

were asked about their test preference and their willingness to pay for that preference,[††] as detailed on page 58.

Responses were received from 76 people (72.4% response rate). Many respondents (41%) gave more discursive answers and did not answer the preference question directly. A majority of these non-responders considered there was not yet enough information to make a choice and were not prepared to speculate about their preferences. A small minority wished to give women the choice. The amount that health commissioners reported they were willing to pay is given in *Figure 25*.

Comments made in the free text area of the questionnaire led us to believe that there might be some ambiguity about certain of the responses from health commissioners. The values given in response to the WTP question may in some cases have been the amount commissioners were prepared to pay for the test rather than the question that was asked about the margins.

Health commissioners were also asked whether they would be prepared to pay for both tests for all women. Only 12 health commissioners (13.3%) indicated that they would.

WTP overview

Table 53 summarises the findings on test preferences and values placed on these test preferences by the four groups asked for this information.

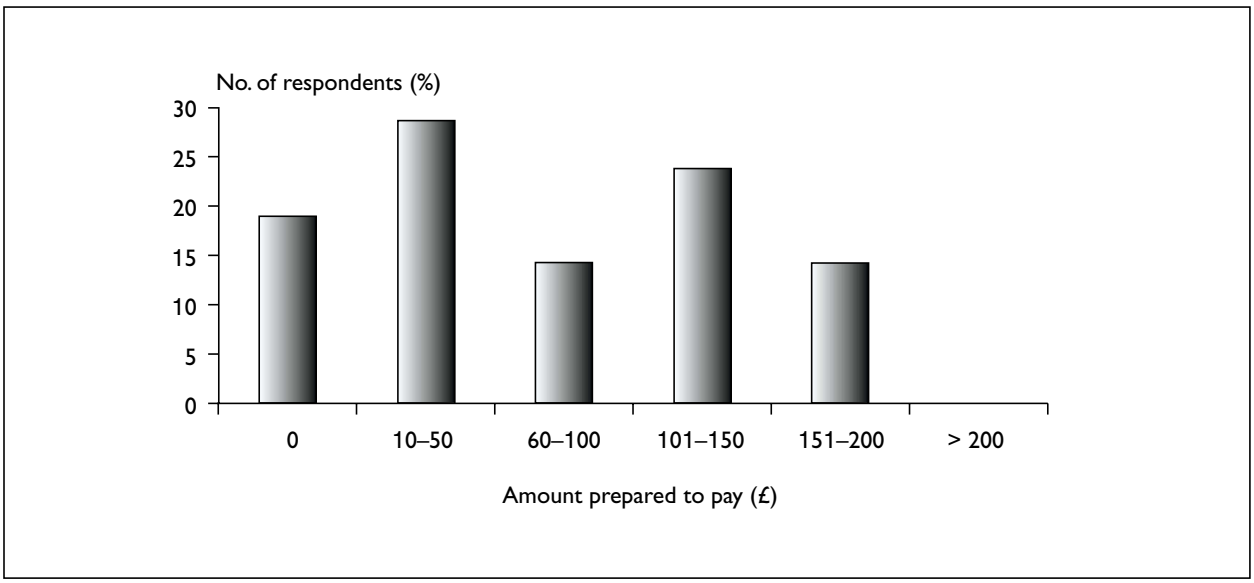


FIGURE 25 *Health commissioners' responses to WTP question*

[††] The question relating to WTP was as follows: "In order to assess the strength of your commissioning preference, overleaf we ask you to hypothesise that a sum would be payable, per test, to the Trust to implement your preference. How much do you think you would be willing to pay, per test, to obtain your given preference?".

Although women reported a statistically significant preference for molecular tests, their strength of preference was not significantly different *ex ante* or *ex post*. Fewer women expressed a preference for molecular tests at the end of the testing period and there was some indication that those who confirmed their original preference for molecular tests strengthened their preference *ex post*.

Women's partners also reported a preference for molecular tests (51%), although the difference was not statistically significant. However, those who wanted karyotyping had a very strong preference for this test, being prepared to pay more than twice the amount for their choice (£125 vs £50).

The general public were unique in identifying a preference for karyotyping (59%). They also reported a strong preference for this test choice (£100 vs £50).

Health commissioners were as likely to report a preference for karyotyping as for molecular tests (i.e. no statistically significant differences between the two). The strength of their preference was the same for both tests (£100).

Overview of economic evaluation techniques

The study adopted three approaches to comparing the costs and benefits of different molecular test regimes relative to traditional karyotyping – cost–utility analysis, cost-effectiveness analysis and cost–benefit analysis. This enabled comparison of the strengths and weaknesses of these approaches in evaluation of prenatal diagnostic tests, although it should be remembered that costs for new technologies can change rapidly as the rate of technology diffusion rises. Thus, the costings used here are necessarily a snapshot of relative costs at the time of the study.

An earlier report on Down screening in this monograph series replaced the concept of cost per case detected by the cost per 'Down syndrome birth avoided' as the common outcome measure to be used in a cost-effectiveness analysis.[8] This study also estimated a 'medical' cost-effectiveness ratio that made allowance for fetal loss as a result of amniocentesis or CVS, as well as including the number of Down syndrome births avoided.[‡‡]

Such outcome measures were considered inappropriate for use in the present study because the diagnostic tests being evaluated are designed to detect a broader range of chromosome abnormalities than just Down syndrome, some of which may be of uncertain clinical significance. Therefore, any analysis that involved the use of 'births avoided' as an outcome measure would by definition ignore the benefits of more detailed information to parents and clinicians on a range of chromosome abnormalities that do not necessarily result in termination of pregnancy; for example, to assist birth and neonatal care planning, and planning of subsequent pregnancies.

In the present study, the first approach used was to conduct a cost-effectiveness analysis to measure technical efficiency using a common output natural to screening systems – the 'cost per case detected'. The study defined 'cases detected' as either (i) all chromosome abnormalities detected by a test without necessarily making any judgement about their clinical significance or (ii) those chromosome abnormalities detected that were judged to be clinically significant. In the case of (i) it is evident that not all 'cases' would necessarily result in a decision to terminate the pregnancy; even in the second instance possibly 50% of 'cases' could be mosaicisms (see page 16, 'Fineberg level 2: molecular test diagnostic accuracy').

As well as a simple cost-effectiveness analysis ratio (cost per case detected), the study also calculated a weighted cost-effectiveness ratio taking into account the rate at which diagnostic information is delivered to women and the information they might be expecting from the test. Such an approach is only possible if it can be assumed that perfect information on all babies is available at or after the birth and thus that the presence of chromosome abnormality is made known to parents eventually.

A second approach that can be adopted in order to value the benefits of molecular tests and compare these to costs is cost–utility analysis. In the present study, utility was expressed in terms of measured health status, or QoL, over the testing period. This was recorded using the EuroQoL (EQ-5D) questionnaire. Thus, a marginal cost per QALY gained was estimated for molecular tests

[‡‡] Medical cost-effectiveness ratio = no unaffected fetuses lost (as a result of amniocentesis/CVS)/no Down syndrome births avoided

compared to karyotyping. However, the study was only able to measure the benefits of the 'new gold standard' testing regime (add-on molecular tests) compared to karyotyping, because of ethical constraints and the stage of development of the molecular tests.

The changes in health status that were measured occurred over a period of weeks, whereas the valuation of health states using EuroQoL EQ-5D is normally based on the assumption that individuals are in a particular health state for a period of 12 months. Therefore, the self-rated health status valuation (VAS) was used in the analysis. Nevertheless the instrument used in this manner appeared sensitive in terms of monitoring changes unidimensionally and on the self-rated health status score. The significance of changes in self-rated health scores was further validated by use of an anxiety-specific measure (Speilberger State–Trait Anxiety Inventory).

A third approach that can be used to value molecular tests is cost–benefit analysis. In the present study, the monetary value (contingent value) that various respondents were willing to pay to obtain their choice of preferred test regime was explored. This approach was only able to compare the constrained choice of molecular tests versus karyotyping (replacement) because of the complexity of presenting information about the relative benefits of combination of tests and the context in which information about the study was delivered.

Each of the approaches adopted was able to address different elements of the benefits that might be provided by molecular tests, either directly or indirectly. All approaches incorporated information on costs of different testing regimes. The information derived from the various approaches is summarised in *Table 54*.

The molecular tests assessed by each of the four techniques in *Table 54* differed. Cost-effectiveness analysis (direct and weighted) was used to assess both Down syndrome only molecular tests and tests for all five common chromosome abnormalities. The cost–utility analysis and the cost–benefit analysis assessed the benefits of the latter only (5-probe FISH or multiplex Q-PCR). Findings for all three approaches are summarised in *Tables 55a* and *55b*.

In this comparison of the results of the evaluative techniques, the two extreme test regimes:

- 'new gold standard' (i.e. molecular tests as add-on to karyotyping)
- 'molecular test for all' (i.e. replacement for karyotyping)

have been used to illustrate the range of results. Cost-effectiveness ratios for specific options between these extremes are summarised in the next chapter.

In *Table 55a* the cost–utility ratios calculated (column 4) are based on self-reported health status, over the testing period and beyond, for women undergoing karyotyping and those receiving a 'new gold standard' test regime with 5-probe FISH or multiplex Q-PCR. Assuming that health status changes would be similar for a testing regime that used a Down only test instead, cost per QALY values can be estimated for these molecular tests as well. *Table 35* indicates that, based on this assumption, Down only FISH would represent the most favourable test to use in a 'new gold standard' test regime (i.e. molecular tests as add-on to karyotyping). However, even in this case the estimated cost per QALY value is much higher than those for other screening programmes, for example US$3190 for diabetic retinopathy screening, around £4000 for breast screening and £8300 for phenylketonuria screening.[75–77]

TABLE 54 *What the different economic analysis techniques used incorporate in assessment of molecular tests*

Analytic technique	Benefit measurement incorporates				
	Cost	Information contents (i.e. case detection)	Time of delivery of result	Patient outcomes (i.e. QoL)	Patient preferences
Cost-effectiveness analysis	✔	✔ directly			
Weighted cost-effectiveness analysis	✔	✔ directly	✔ directly		
Cost–utility analysis	✔	✔ indirectly	✔ indirectly	✔ directly	
Cost–benefit analysis	✔	✔ indirectly	✔ indirectly	✔ indirectly	✔ directly

TABLE 55a *Results of different economic evaluation approaches based on laboratory with throughput of 1000 samples of molecular tests per annum*

1	2	3	4	5	6	7	8
Test regime	**Test configuration**	**Test type**	**Cost–utility analysis[a] Cost per QALY**	**Detection of all chromosome abnormalities**		**Detection of Down syndrome only**	
				Unweighted cost-effectiveness analysis[a] Cost per case (% difference from karyotyping)	**Weighted cost-effectiveness analysis[b] Cost per case (% difference from karyotyping)**	**Unweighted cost-effectiveness analysis[c] Cost per case (% difference from karyotyping)**	**Weighted cost-effectiveness analysis[d] Cost per case (% difference from karyotyping)**
'New gold standard'	Down syndrome only	FISH	£23,542	£3,183 (+47%)	£1,366 (–37%)	£8,311 (+47%)	£902 (–84%)
		Q-PCR	£24,670	£3,232 (+49%)	£1,467 (–32%)	£8,439 (+49%)	£916 (–84%)
	All common	5-probe FISH	£41,939	£3,979 (+84%)	£960 (–56%)	£10,388 (+84%)	£1,128 (–80%)
		Multiplex Q-PCR	£26,514	£3,312 (+53%)	£908 (–58%)	£8,647 (+53%)	£939 (–83%)
Molecular test replacing karyotyping ('one for all')	Down syndrome only	FISH		£2,657 (+23%)	£9,826 (+354%)	£2,657 (–53%)	£288 (–95%)
		Q-PCR		£2,784 (+29%)	£11,096 (+412%)	£2,784 (–51%)	£302 (–95%)
	All common	5-probe FISH		£2,768 (+28%)	£4,250 (+96%)	£4,733 (–16%)	£514 (–91%)
		Multiplex Q-PCR		£1,750 (–19%)	£3,347 (+55%)	£2,992 (–47%)	£325 (–94%)
Status quo	Karyotyping			£2,166 (0%)	£2,166 (0%)	£5,654 (0%)	£5,654 (0%)

[a] *Unweighted cost-effectiveness ratio = cost per case detected, including all cases of chromosome abnormality whether clinically significant or not*
[b] *Weighted cost-effectiveness ratio = cost per case detected, including all cases and allowing for the relative time for which women remain in states of uncertainty*
[c] *Unweighted cost-effectiveness ratio = cost per case detected for Down syndrome only*
[d] *Weighted cost-effectiveness ratio = cost per case detected for Down syndrome only, allowing for the relative time for which women remain in states of uncertainty about Down syndrome*
[e] *Cost per QALY, based on self-assessment of health status;VAS*

TABLE 55b *Results of different economic evaluation approaches, cont'd*

1	2	3	4	5	6	7	8	9
Test regime	**Test configuration**	**Test type**	**Cost per woman tested**	**Women: pre-test WTP (£) (% expressing this preference)**	**Women: post-test WTP (£) (% expressing this preference)**	**Women's partners WTP (£) (% expressing this preference)**	**Health commissioners WTP (£) (% expressing this preference)**	**General public WTP (£) (% expressing this preference)**
'New gold standard'	Down only	FISH Q-PCR	£131.82 £133.84					
	All common	5-probe FISH Multiplex Q-PCR	£164.75 £137.14					
Molecular for all replacing karyotyping	Down only	FISH Q-PCR	£42.14 £44.16					
	All common	5-probe FISH Multiplex Q-PCR	£75.07 £47.46	£70 (67%) £70 (67%)	£100 (51%) £100 (51%)	£50 (51%) £50 (51%)	£100 (29%) £100 (29%)	£50 (38%) £50 (38%)
Status quo	Karyotyping		£89.68	£95 (32%)	£100 (42%)	£125 (43%)	£100 (29%)	£100 (59%)

The cost–utility analysis does not directly incorporate differences in test performance. However, simple cost-effectiveness ratios (*Table 55a*, column 5) can be estimated that take these factors into account and calculates a cost per case detected, based on all cases detected by a test. Comparison of these cost-effectiveness ratios indicates that all 'new gold standard' test regimes (i.e. add-on molecular tests) are less cost-effective than the status quo (karyotyping). This conclusion would appear to be similar to that reached using cost–utility analysis. Cost-effectiveness analysis also indicates that 'molecular test for all' as a replacement for karyotyping is also generally less cost-effective than karyotyping, with the exception of multiplex Q-PCR.

If unweighted cost-effectiveness ratios are examined, the 'new gold standard' regime for all women is judged to be the least cost-effective option compared to karyotyping, and only replacement of karyotyping by the multiplex Q-PCR molecular test for all women would be more cost-effective.

However, if the cost-effectiveness ratios are weighted to allow for the uncertainty experienced by women waiting for a result,[§§] because information is received soon after amniocentesis via the molecular test and also more complete information is received with the karyotyping result, the 'new gold standard' emerges as a more cost-effective option than karyotyping for all women. However, using weighted cost-effectiveness ratios, a molecular test regime used as a replacement for karyotyping now appears to be less cost-effective than karyotyping, even for the multiplex Q-PCR test. In addition, use of molecular tests as a replacement for karyotyping will lead to some chromosome abnormalities, which previously would have been detected prenatally, not being detected until later or remaining undetected. It is estimated that every year in the UK this would result in approximately 91 clinically significant chromosome abnormalities that would not be detected in the prenatal period. Some of these pregnancies would not go to term, and there would be variable effects on the fetus or child associated with the remainder.

All the analyses so far have assumed that diagnosis is designed to detect all chromosome abnormalities present. If it is assumed that the women being tested expect, and want, only a result for Down syndrome, then both the simple and weighted cost-effectiveness ratios can be re-expressed to take this into account. In this case, the unweighted molecular test regimes as a replacement for karyotyping are now estimated to be more cost-effective than both karyotyping and the 'new gold standard' (*Table 55a*, column 7). Applying weighting, all molecular test regimes are now judged to be more cost-effective than karyotyping because a complete result (for Down syndrome) is reported more quickly even when molecular tests replace karyotyping. In this situation, although both FISH and Q-PCR are judged cost-effective, the exact cost-effectiveness ratio would be highly throughput-dependent, with FISH likely to be more cost-effective for smaller laboratories (*Figure 8*).

These results demonstrate the extreme sensitivity of an approach such as cost-effectiveness analysis to the outcome measure chosen. Using the simple measure of the number of cases detected takes into account only one of the features of molecular tests. Although unproven in terms of the assumptions made, the weighted cost-effectiveness analysis demonstrates the point that the results can be very different if slightly different assumptions are used. It should be emphasised, however, that the values presented in the weighted cost-effectiveness analysis should be used with extreme caution, until further research can establish the nature of the functions assumed in this analysis to be linear.

In addition, comparisons of the technology discussed in this report with other potential healthcare investments can only be made across the limited number of interventions aimed at achieving the same outcome. Thus the outcome, cost per case detected for prenatal diagnosis of chromosome abnormalities, can be used for comparison with other screening programmes. However, comparison cannot be made within the cost-effectiveness framework with other healthcare interventions competing for investment of resources (allocative efficiency) and the cost–utility analysis will inform in this domain.

The cost–utility and cost-effectiveness analyses concentrated on the impact of different tests on the health status of women or on the numbers of chromosome abnormalities detected respectively, and the extended cost-effectiveness analysis also considered the points at which abnormalities were identified. Neither approach considered preferences, which were addressed through a cost–

[§§] As has been commented earlier, 97% of women receive reassurance from a negative result.

benefit analysis. *Table 55b* provides information on the preferences reported by different stakeholders for molecular tests versus karyotyping, where the choice was constrained to either a karyotyping or molecular test, and the strength of preference for the choice made. The strength of preference was expressed, within a WTP framework, as how much individuals are prepared to pay to obtain the test of their choice. From the responses it is clear that a majority of women preferred molecular tests. However, partners and the general public[11] stated a preference for karyotyping, and also expressed a stronger preference for their choice than women. The amount that individuals were prepared to pay for their choice, if this choice was karyotyping, was largely similar. All groups were prepared to pay more than the marginal cost of providing a test for chromosome abnormality.

Because molecular tests are less expensive than karyotyping, there is no question of the costs associated with the choice made by women not being met. However, the incremental cost of providing molecular tests as an add-on to karyotyping (i.e. ranging from £42.14 for Down only Q-PCR to £57.07 for 5-probe FISH), while low for Q-PCR when compared to the strength of preference reported, is much higher for FISH. When strength of preferences for FISH are compared, some of the values are less than the incremental cost of providing the molecular test as an add-on to karyotyping.

[11] The general public conclusion is based on a sample that was not representative of population in the area of the survey, as respondents tended to be both older and wealthier than the mean.

Chapter 9

Summary and conclusions

Service delivery

The impact that the introduction of molecular tests is likely to have on various aspects of prenatal diagnostic service delivery can be summarised as follows.

Impact on screening programmes

Screening for chromosome abnormalities in pregnancy is a two-stage process. The overall effectiveness of the combined stages:

- primary case selection
- testing of amniocytes or chorionic villi samples

in terms of abnormal cases detected is dependent on effective primary case selection for diagnostic testing. Case selection is by assessment of risk based on parameters such as familial risk, maternal age, serum testing results and ultrasound observations, which are then summated into an overall risk score. There are local variations within the UK in this primary case selection process, with observed differences in detection rates, such as those recorded for Northern Ireland and the West Midlands in the present study (see *Tables 9* and *10*), arising in part from these variations in case selection.

The widespread introduction of molecular tests will offer a useful opportunity to refine testing protocols at a local level at the same pace as case selection protocols are developed for a National Down Syndrome Screening Programme. However, this report makes it clear that the process of refining testing protocols will only be successful if multidisciplinary interest groups initiate effective dialogue based on this and other recent reports.[8,13]

The widespread introduction of molecular tests should also encourage the development of strategies that tailor the type of diagnostic test offered to the risk identified. Thus, if a case is selected because of a risk assessment that indicates an increased risk of Down syndrome, it may be appropriate as well as more efficient to offer a test that detects Down syndrome only. The possibility of such refinement of test selection, following the introduction of molecular tests, is explored further in appendix 9, where an illustration of a test selection pathway can be found.

There is no evidence from the present study, however, that molecular tests will lead to increased levels of prenatal diagnostic testing (i.e. supply-induced demand) because the decision to test is based on estimation of harm versus benefit, namely miscarriage risk versus probability of chromosome abnormality. Miscarriage risk will not be modified until acquisition of fetal cells no longer requires an invasive procedure such as amniocentesis. Once fetal cells can be obtained routinely from the maternal blood, however, further refinement of primary case selection and appropriate tailoring of testing regimes will become crucial. Under these conditions, also, demand for prenatal testing for chromosome abnormalities might increase sharply as may, therefore, the use of molecular tests.[78–81]

Impact on pre-test counselling

Because molecular tests, by design, diagnose a specific, and more limited, range of abnormalities than karyotyping, the recommendations of the Advisory Committee on Genetic Testing[13] are particularly pertinent for the introduction of these tests, namely:

- At each pregnancy, bearing in mind advances in technology and knowledge, women should be offered information on prenatal tests appropriate to their individual risk.
- Information should enable the women to understand the nature of the test, its scope and limitations, and the accuracy, significance and use of the result.

Clearly, part of pre-test counselling will also involve managing patients' expectations about the duration of wait for a test result.

It does not appear, based on the views of obstetricians and midwives recorded in the study, that the introduction of molecular tests will place additional burdens on the current pre-test counselling process. The majority of obstetricians in the national survey (57%) reported that they considered that there would be no extra burden

and only one in four (25%) considered that the way in which test results are currently administered by their Trust would have to change. Similarly, most midwives (67%) reported that molecular tests would not place an additional burden on counselling services, and only 14% thought that result-giving would be affected. The main conclusion drawn from both surveys was that the introduction of molecular tests is not expected to adversely affect counselling or change the result-giving process. However, as noted on page 59, assessment of the full impact of false-negatives, if molecular tests are to replace some karyotyping tests, is outside the scope of this study and there would appear to be a need for further research in this area.

Impact on antenatal services

Pressures to fully realise the benefits from a faster turnaround time for molecular tests in the laboratory will have important implications for local antenatal clinics and community services. Any delays in clinic staff delivering results could significantly add to test result times for women, thereby reducing the potential gains provided by the more rapid laboratory processing.

It is clear from the national laboratory survey that, for abnormal results, strenuous efforts are currently made to contact women more rapidly than for normal results.* Prospectively, all patients must anticipate a positive result as each individual has already been told they are undergoing amniocentesis after having been identified as at higher risk. Perversely, within the current system, therefore, those who have least to be anxious about wait longest. Because the benefit of reduced anxiety is the major positive outcome for rapid molecular tests, clinics and laboratories must jointly explore new mechanisms for the timely transmission of results. Such mechanisms must reflect the shortened test turnaround time and also be equitable for all those tested, including the 97% of women with a negative result as well as the 3% with a positive result. Links already exist for other diagnostic services such as biochemistry and pathology to transmit results via NHSNet or other dedicated information systems directly to the clinician. Forming such a link between cytogenetic and antenatal services should be given priority at the same time as the introduction of new molecular test processes in the laboratory. Where the referring clinician is happy for the laboratory to report negative results directly to the patient, the use of secure modern technologies may also be considered. (Responses to the UK laboratory survey indicated, however, that only two laboratories report negative test results directly to women by post.)

Impact on laboratories

Three-quarters (22/30) of UK laboratories surveyed reported that they were not currently using molecular tests for routine prenatal testing. However, over half of these (12/22) were in the process of assessing FISH and 3/22 were assessing Q-PCR. It would appear, therefore, that overall laboratories were preparing to implement FISH at the time of the survey in 1998. In addition, from free text supplied by laboratory directors with the surveys, it was suggested that FISH is viewed as a test that might be more suited to the "culture and skills" of cytogeneticists. However, when asked for a preferred molecular test less than half (47%) of those not using molecular methods indicated a definite preference for FISH. Laboratories did not anticipate problems with training of staff to undertake either molecular test, although only 64% would train in-house for Q-PCR compared with 94% for FISH. Recruitment difficulties were explored, but laboratory directors indicated that problems tended to be those of recruitment of clinical scientists in general or of geographical location, rather than linked to a specific molecular test.

The study measured the direct laboratory cost of different molecular tests. This analysis indicates that for a laboratory with an annual throughput of 1000 specimens, and appropriate skill-mix, molecular tests are substantially less expensive than karyotyping: £90 direct test cost for karyotyping versus £75 and £42 for 5-probe and Down only FISH respectively, and £47.50 and £44 for multiplex and Down only Q-PCR.

A further conclusion, based on cost grounds, is that Q-PCR tests are particularly suited to medium or high throughput laboratories (with throughputs above 1100 specimens per annum for multiplex tests); and that FISH would be favoured in lower throughput laboratories (fewer than 450 specimens per annum (see *Figure 7*). Approximately 23% of UK laboratories reported throughputs below 450 samples per annum and 53% above 1100.

* Normal results are, for example, sent by letter to the consultant whereas abnormal results are phoned through to the antenatal clinic or community midwife.

If prenatal testing were to be centralised further in the UK, either through amalgamation of current services or by partitioning of prenatal testing, the resulting increase in the number of higher throughput laboratories would favour the choice of Q-PCR on cost grounds.

Evaluation framework

The framework used for the assessment of molecular tests is based on an adapted form of the hierarchy for diagnostic tests proposed by Fineberg in 1977 (appendix 10). The study's conclusions on each of these levels are summarised below.

Technical capacity and diagnostic accuracy

The first two levels of assessment addressed the technical capacity and diagnostic accuracy of the tests. Molecular tests were first assessed in the laboratory and then evaluated in the service setting. The conclusions drawn at the end of these two stages of the study were that both FISH and Q-PCR are reliable and precise for detection of Down syndrome. If diagnosis of Down syndrome only is being considered, molecular tests are a good substitute for karyotyping in terms of both cost and production of rapid results. One of the features of these molecular tests, however, is that they are not designed to detect the full range of chromosome abnormalities that karyotyping can detect. Thus, direct comparison of the diagnostic performance of molecular tests is complex and requires careful interpretation (appendix 10). Throughout the assessment of test performance, therefore, two perspectives were used.

The first perspective considered the performance of molecular tests in terms of the abnormalities the tests are capable of detecting – their absolute performance in comparison with karyotyping for these same abnormalities. From this perspective, all molecular tests were considered to have an acceptable sensitivity and specificity to use for patient testing. This conclusion has the caveat, however, that rates of inconclusive results were higher for both types of molecular tests (about 7–9%) than for karyotyping (0.2%). Comparable levels of inconclusive results have been reported for FISH in other recent large-scale studies.[61] In addition, *Table 8* shows that Q-PCR exhibited a slightly lower absolute sensitivity and specificity than FISH, 0.9565 and 0.9997 respectively. Q-PCR also appeared to be sensitive to maternal cell contamination of the amniotic fluid sample.

As a replacement for karyotyping, therefore, it is concluded that back-up alternative tests may be required for a small proportion of samples that produce inconclusive test results from the first molecular test. The implications of this possibility are two-fold. First the attraction of molecular tests requiring smaller samples of amniotic fluid than for karyotyping must be set against the need for sufficient reserve to repeat the test. Second, the counselling process should inform patients of the possible delay that second testing will introduce.

A second perspective for the technical assessment was in terms of all the abnormalities that karyotyping can detect, and therefore considered the relative performance of molecular tests against this absolute (the current gold standard test). In terms of their relative performance it is estimated that, for molecular tests designed to detect the five most common chromosome abnormalities, if these tests were to completely replace karyotyping then approximately 4.2 clinically significant chromosome abnormalities would remain undetected for every 1000 women tested. This figure is calculated from estimated rates of prevalence, based on audits in the two test regions, Northern Ireland and the West Midlands, and should therefore be treated with some caution. The cases defined as clinically significant included approximately 50% that were mosaicisms, with a different profile in terms of the likely effect on the child (see page 17, 'Clinical significance'). The resulting 2.1–4.2 clinically significant undetected chromosome abnormalities per 1000 women tested can be compared with the current fetal loss rate of 9 per 1000 associated with mid-trimester amniocentesis or the estimated loss rate of 36 per 1000 for early amniocentesis or transcervical CVS.[8] Therefore, the risk of missed chromosome abnormalities with molecular tests is comparable to other risks currently associated with the testing process; although the latter will be significantly reduced once fetal cells can be obtained through other routes (i.e. maternal blood).

Finally, molecular tests might be perceived to offer an advantage over karyotyping in situations where relatively little amniotic fluid is available. In both early and late pregnancy there may be difficulties in acquiring adequate sized samples of fetal cells for karyotyping. However, it should be noted that there were more failed molecular tests for early first trimester samples, and those taken later in the second trimester, than for mid second trimester samples (*Table 5*). Conversely, larger sample sizes would be needed in order to routinely complete karyotyping and a molecular test successfully, and

local protocols should be designed to alert clinicians to the relationship between range of analyses requested and the sample size required. In practice, difficulties in obtaining large enough samples to perform both tests were rarely encountered in the present study.

Diagnostic impact

The third level of assessment of molecular tests considers the extent to which the new test will replace other diagnostic tests or procedures. Stakeholder groups were consulted and their preferences for molecular test use explored. For obstetricians, when the option was a constrained choice between either a molecular test or karyotyping, only 9% in the national survey reported they would choose karyotyping, while 40% would select a molecular test (*Tables 29* and *30*). The remainder were either undecided (27%) or would give women the choice (24%). If molecular tests were to replace karyotyping, the preferred test configuration selected by obstetricians was a molecular test for the five common abnormalities (57%). When asked about molecular tests as a replacement for karyotyping, only 2% expressed a preference for a Down only test, although when asked to review specific cases retrospectively it was judged by the clinician responsible that a Down only molecular test would have been selected in 21% of cases for which questionnaires were completed.

Finally, when asked to state their preferences given an unconstrained choice, 57% of obstetricians expressed a preference for molecular tests, with only 15% reporting they would choose the 'new gold standard' regime (both tests for all, or a majority, of women; *Table 28*). Thus it may be difficult to identify a single molecular testing regime that will satisfy all clinicians, although clearly molecular tests are preferred over karyotyping, with relatively few obstetricians expressing a preference for both tests ('new gold standard'). Similarly, only 13% of health commissioners indicated that they would be willing to pay for both tests.

When women themselves were asked to state their preference (for a constrained option), two-thirds (67%) expressed an *ex ante* preference for molecular tests and one-third (32%) for karyotyping.

Patient outcome and cost-effectiveness

The final two assessment levels address the impact of a test on patient outcomes and its cost-effectiveness. The study measured the extent to which use of molecular tests contributed to improved health or reduced anxiety for women undergoing testing, and also whether use of molecular tests could improve the cost-effectiveness of diagnosis compared to the current test, karyotyping.

A generic QoL instrument (EuroQoL EQ-5D) was used to measure changes in health status during the testing process. The findings were then used in a cost–utility analysis to aid comparison with alternative healthcare interventions. Although the reliability and validity of the QoL instrument when used for short-term changes in health status is still under investigation, the study was able to demonstrate significant changes in self-rated health status scores linked to more rapid test results. This was further validated by use of an anxiety-specific measure (Speilberger State–Trait Anxiety Inventory), which demonstrated similar significant differences between women who received a rapid molecular test and those who were tested using karyotyping.

The QoL instrument clearly highlighted a gain in health status experienced by women receiving a molecular test result compared to those waiting for a karyotyping result. However, the cost–utility analysis also demonstrated that the cost per QALY gained through the use of molecular tests as a routine add-on to karyotyping (the 'new gold standard') may not represent good value for money when compared with other uses for these resources. The marginal cost per QALY gained is in the range £23,542 to £26,514 for Down only FISH and Q-PCR or multiplex Q-PCR. This is close to the informal thresholds of acceptability for bodies such as the National Institute for Clinical Excellence. The estimated value for use of 5-probe FISH was much higher, at £41,939 per QALY.

In laboratories processing 1000 samples per annum, therefore, the cost–utility analysis indicates that the use of any molecular test configuration (with the exception of 5-probe FISH) in addition to karyotyping ('new gold standard') would be judged borderline as an acceptable use of resources. This conclusion assumes that the benefit of more rapid positive results is the same as that of negative results.

Inherent weaknesses in the cost–utility analysis include its dependence on the valuation of impact on QoL in the short term and use and extrapolation of self-rated health status between measurement points. However, measurements of anxiety using a different instrument (the Speilberger State–Trait Anxiety Inventory)

supported these findings, by demonstrating a similar pattern of impact on this element of QoL. It should be noted that the cost–utility analysis was not able to inform on the introduction of molecular tests as a replacement for karyotyping because it was not possible ethically to trial molecular tests as a sole replacement test. However, as long as women require and expect a test for Down syndrome only, or for the most common abnormalities, replacement molecular tests should represent better value because the test is less expensive than karyotyping and confers a higher utility arising from a faster result.

As well as measuring cost–utility, the study also undertook a cost-effectiveness analysis, comparing molecular tests with karyotyping. Cost-effectiveness was estimated in the first instance in terms of the cost per case detected. *Table 53* summarised the resulting cost-effectiveness ratios for the broad range of testing regimes suggested by clinicians.

Table 56 lists, for each test regime, the molecular test (type and configuration) calculated to have the most favourable cost-effectiveness ratio. All the cost-effectiveness ratios in this table are calculated at a laboratory throughput of 1000 molecular tests per annum. The table also indicates the level of undetected clinically significant chromosome abnormalities for each testing regime although, as has been pointed out previously, the definition of clinically significant cases included chromosome abnormalities such as mosaicisms, which accounted for 50% of such cases in our audits.

Table 56 demonstrates there is no single molecular test that has a more favourable cost-effectiveness ratio for all the different regimes that include molecular tests. The molecular test regimes with a lower cost per case detected than karyotyping appear to optimise cost-effectiveness ratios at the expense of the number of chromosome abnormalities detected. It should also be noted that Q-PCR, which emerges most frequently as the most cost-effective test at a throughput of 1000 cases per annum, is highly sensitive to throughput.

Table 57 shows that at throughputs below approximately 450 cases per annum, FISH will become the favoured molecular test in all the regimes where Q-PCR was the more cost-effective option in *Table 56*. Conversely, for higher throughputs (above 1200 samples per annum), FISH would no longer be the test for the two test regimes in which it was the more cost-effective option in *Table 56,* and Q-PCR would be the molecular test with the best cost-effectiveness ratio for all regimes.

Table 57 shows that a further important effect of decreasing throughput is to drive all molecular test cost-effectiveness ratios up beyond that of the status quo, making karyotyping the most cost-effective option. Therefore, in situations where FISH is predominant, karyotyping is the more cost-effective test option. However, karyotyping delivers a slower result even though the information provided is more comprehensive. Therefore, some of the

TABLE 56 *Optimum molecular test (based on cost-effectiveness ratios) for various test regimes at a throughput of 1000 cases per annum*

Option		Test regime	Most cost-effective molecular test (type and configuration)	Unweighted cost-effectiveness ratio (variation from status quo)	Undetectable clinically significant cases per 1000 tests[a]
1		'New gold standard'	Down only FISH	£3183 (+47%)	0
2	A	Karyotyping for all (status quo)		£2166 (0%)	0
	B	'Molecular test for all'	Multiplex Q-PCR	£1750 (–19%)	4.2
3		Molecular for all, karyotyping for high-risk cases	Multiplex Q-PCR	£1665 (–23%)	2.5
4		Karyotyping for all, molecular for high-risk cases	Down only FISH	£2233 (+3%)	0
5		Parental choice, 'new gold standard' for high-risk cases[b]	Multiplex Q-PCR	£2270 (+5%)	1.5

[a] *See page 65*
[b] *See page 57: 60% would choose molecular tests, 33% choose karyotyping, 7% receive 'new gold standard'*

TABLE 57 *Optimum molecular test (based on cost-effectiveness ratios) for various test regimes when throughput falls to 450 cases per annum*

Option		Test regime	Most cost-effective molecular test (type and configuration)	Unweighted cost-effectiveness ratio (variation from status quo)	Undetectable clinically significant cases per 1000 tests[a]
1		'New gold standard'	Down only FISH	£3183 (+47%)	0
2	A	Karyotyping for all (status quo)		£2166 (0%)	0
	B	'Molecular test for all'	5-probe FISH	£2768 (+28%)	4.2
3		Molecular for all, karyotyping for high-risk cases	5-probe FISH	£2527 (+17%)	2.5
4		Karyotyping for all, molecular for high-risk cases	Down only FISH	£2233 (+3%)	0
5		Parental choice, 'new gold standard' for high-risk cases[b]	5-probe FISH	£2890 (+33%)	1.5

[a] *See page 65*
[b] *See page 57: 60% would choose molecular tests, 33% choose karyotyping, 7% receive 'new gold standard' (i.e. both)*

options in *Table 57* (and also *Table 56*), which give more information earlier, may be preferred both by clinicians and by women waiting for their test result.

Equally importantly, the cost-effectiveness ratio for karyotyping (£2166 per case detected) is based on all cases that this test can detect, and includes a number of non-clinically significant cases. If only clinically significant cases are considered, then the ratio for karyotyping rises from £2166 to £2868 per case detected, because the number of cases included falls. Based on the values in *Tables 56* and *57*, this would make all molecular test regimes more cost-effective than karyotyping except for the following: (i) parental choice at a throughput of 450 samples or below per annum and (ii) the 'new gold standard' regime at throughputs of 1000 or fewer samples per annum. Were clinically significant cases to be limited to Down syndrome only, the cost per case detected for karyotyping would rise to £5654.

It is possible to extend the simple cost-effectiveness ratios presented above to include weighted values (weighted to make allowance for the degree of uncertainty remaining after a test result and the time spent in different states of uncertainty). If such weighted ratios are calculated, then the pattern of test option with the most favourable cost-effectiveness ratio changes. The results of such an analysis are presented in *Table 58* for a throughput of approximately 450 cases per annum (comparable to *Table 57* where Q-PCR is the more expensive molecular test). This table shows that, at low throughputs, the most cost-effective test regime is now calculated to be the 'new gold standard' regime (rather than karyotyping as in *Table 57*). The 'molecular test for all' testing regime is now calculated to be even less cost-effective than before.

TABLE 58 *Weighted cost-effectiveness ratios for best alternative to Q-PCR where throughput is 450 cases per annum*

Option[a]		Test regime	Most cost-effective molecular test (type and configuration)	Unweighted cost-effectiveness ratio (variation from status quo)	Undetectable clinically significant cases per 1000 tests[a]
1		'New gold standard'	Down only FISH	£985 (–55%)	0
2	A	Karyotyping for all		£2166 (0%)	0
	B	'Molecular test for all'	5-probe FISH	£4491 (+107%)	4.2
5		Parental choice, 'new gold standard' for high-risk cases[b]			

Although the combined molecular testing and karyotyping regime is the most cost-effective in terms of weighted ratios, the overall cost of testing would rise by 53% in these smaller laboratories (*Table 37*).

The weighted cost-effectiveness ratios above are presented only for this one example because they are intended to be illustrative cases. It should also be borne in mind that the form of weighting used in the weighted cost-effectiveness ratios has not yet been validated. This extension of the analysis has been performed to demonstrate that it should be appreciated that the conclusions drawn from cost-effectiveness analysis relating to this particular diagnostic test are very sensitive to the outcome measures chosen.

In conclusion, for the 'new gold standard' regime, based on a simple cost-effectiveness analysis (using the cost per case detected as the comparator), it is calculated that no molecular test can make the 'new gold standard' regime more cost-effective than karyotyping at a throughput of 1000 tests per annum. Even if this analysis only considers the clinically significant cases diagnosed by karyotyping, the situation remains the same. Under these conditions, Down only molecular tests would be preferred as an add-on to karyotyping, but the regime is not judged to be cost-effective. However, if the cost-effectiveness ratios are weighted then the conclusions that emerge are closer to those drawn by the cost–utility analysis. Using this approach karyotyping plus any molecular test (including 5-probe FISH) is identified as being more cost-effective than karyotyping on its own. Also, multiplex Q-PCR emerges as the optimal test to use, rather than one of the Down only molecular tests identified by the cost–utility analysis.

If the 'new gold standard' regime is considered for smaller laboratories (below 1000 samples per annum), the use of molecular tests as an add-on will become increasingly less cost-effective. For larger laboratories the reverse would be the case, with add-on molecular tests becoming more favourable with increased size. If both tests are performed on a sample there may be some savings on a simple sum of the costs of the two independent tests. Balanced against this reasonably positive picture is the fact that only 15% of clinicians reported that they would be likely to order both tests for a woman, and only 13% of health commissioners said that they would be willing to pay for both tests.

In contrast, the possible introduction of molecular tests as a replacement for karyotyping produces a different picture. At an annual throughput of 1000 samples, the simple cost-effectiveness analysis indicates that only multiplex Q-PCR can be demonstrated to be cost-effective. If the analysis is refined to include only clinically significant cases as the comparator, then all four molecular tests are judged to be more cost-effective than karyotyping. However, the use of weighted cost-effectiveness ratios would appear to indicate that none of the molecular tests (even multiplex Q-PCR) are cost-effective as a replacement for karyotyping; this conclusion does not change if only clinically significant cases are considered for karyotyping. Furthermore, the underlying assumption in this instance is that all chromosome abnormalities that are not identified by molecular tests will be disclosed at birth.

The study demonstrates that if molecular tests were introduced as a replacement, significant savings could be realised, especially if services are configured to operate at higher throughputs (see *Figure 7*). At the same time, it is estimated that approximately 2.1–4.2 clinically significant, less common chromosome abnormalities would be missed per 1000 cases tested if molecular tests were to replace karyotyping. However, careful case selection to identify those women undergoing genetic testing outside a Down screening programme who are at higher risk of the less common abnormalities, and provision of full karyotyping for these women, should keep undetected clinically significant abnormalities to a minimum.

Finally, the study also measured preferences and willingness to pay for the different test alternatives. Women reported a statistically significant preference for molecular tests *ex ante* with a median WTP of £70 for their choice of test. Those preferring karyotyping reported a stronger preference (£95 WTP, $p = 0.001$). Both of these values would approximately cover the incremental cost of providing the requisite test in a laboratory with a 1000 annual throughput (multiplex Q-PCR at £47/5-probe FISH at £75, and karyotyping at £90). Women's partners similarly expressed a preference for molecular tests *ex ante*, but the strength of preference was much higher for karyotyping (£125) than for molecular tests (£50).

The picture that emerges from the cost–benefit analysis, therefore, demonstrates that the amount women are willing to pay *ex ante* for their preferred test choice is close to the incremental cost of

providing the preferred test, whether it is a molecular test or karyotyping. Women's partners are willing to pay even more for karyotyping, far above the incremental cost of performing this test, and their recorded WTP value for molecular tests is also slightly above their incremental cost (except for the case of 5-probe FISH). In smaller laboratories, the costs of molecular tests would be higher and therefore less favourable when compared to WTP values.

In conclusion, the picture that emerges overall suggests that in a laboratory processing 1000 samples per annum, the 'new gold standard' regime may be potentially worthwhile in terms of its impact on QoL, parents' WTP and weighted cost-effectiveness ratios but at an overall increased cost of testing of around 50%. It would not appear worthwhile, however, based on simple cost-effectiveness ratios. Furthermore, only 15% of clinicians reported that they would be likely to order both tests for a woman, and only 13% of health commissioners said that they would be willing to pay for both tests. There would appear, therefore, to be strong reasons for offering molecular tests in addition to karyotyping where women are judged at risk of chromosome abnormalities other than Down syndrome or where the anxiety caused by the wait for karyotyping is an important outcome for the woman and her partner. In addition, to offer molecular tests within a Down screening programme might help rationalise counselling and improve outcomes. Nevertheless, some of the issues of utility gains from the various test regimes and the preferences recorded would support the view that a more flexible protocol may be the best way forward. Greater equity may be achieved by selecting high-risk cases for the 'new gold standard' regime and offering other women the choice of a molecular test or karyotyping.

Conclusions

Conclusions can be drawn from likely options for widespread introduction of molecular tests in the UK. The main options considered were molecular tests as an add-on to karyotyping, as a replacement for karyotyping and a hybrid model somewhere between these two extremes. The conclusions within each of these options will be presented in turn.

Molecular tests as an add-on to karyotyping

- A molecular test giving a rapid result for most common conditions followed by the additional information about rarer conditions conveyed by the karyotyping result can be regarded as a 'new gold standard' test for chromosome abnormalities, combining maximum reduction in anxiety with maximum information.
- Results received 2 or 3 days after testing have a significant impact on QoL in terms of reduction in anxiety, compared with that experienced during the longer wait for karyotyping results.
- Within a National Screening Programme for Down Syndrome, where risk assessment is for Down syndrome only, the 'new gold standard' regime is not a cost-effective option.
- For the 'new gold standard' regime, using molecular tests for the five most common chromosome abnormalities, the cost per QALY gained for 96% of women who will receive a negative test result is approximately £25,000 (with the exception of FISH at £42,000) in laboratories with the median UK throughput.
- If the value for money of the 'new gold standard' regime is explored, using cost per case detected, karyotyping is a more cost-effective option than 'new gold standard', which would also increase UK test costs by 47% (using best molecular option, Down only FISH). However, if cost-effectiveness ratios are modified to include weightings for information content of test and anxiety reduction, then the 'new gold standard' appears to be the more cost-effective option.
- If throughput in a laboratory falls below 450 amniotic fluid samples per annum then FISH will always be the molecular test of choice but may not be the most cost-effective option compared to karyotyping.
- Stakeholders (parents, clinicians and commissioners) valuations of tests are broadly similar to the incremental cost of a test. This value is measured using WTP for test of choice (contingent valuation).
- Surveys reveal that the 'new gold standard' is the preferred option for only 15% of obstetricians and 13% of health commissioners.

Molecular tests as a replacement for karyotyping

- Based on all chromosome abnormalities detected, all configurations of molecular tests are less cost-effective than karyotyping, except multiplex Q-PCR, which costs 19% less per case detected in laboratories with a throughput of 1000 samples per annum. This is because molecular tests do not detect all the chromosome abnormalities detectable by karyotyping. However, if clinically significant cases only are compared (like with like) then all molecular test configurations are demonstrated to be more

cost-effective than karyotyping, based on the cost per case detected.
- If molecular tests were to replace karyotyping, between 2.1 and 4.2 clinically significant abnormalities not relating to trisomies 21, 18, 13 or numerical sex chromosome abnormalities would remain undetected per 1000 samples tested.
- As part of a National Down Screening Programme it would be most cost-effective to configure molecular tests to detect Down syndrome only. In this case both Q-PCR and FISH are more cost-effective than karyotyping, with FISH preferred for smaller laboratories (< 450) and Q-PCR for larger laboratories (> 1100). Use of even larger laboratories will produce significant further economies of scale for Q-PCR. This option should be debated as it raises complex issues with respect to counselling of patients.
- If cost-effectiveness ratios are modified to include weightings for information content of test and anxiety reduction then no molecular test configuration is judged to be a cost-effective replacement for karyotyping.
- This study did not pursue cost–utility analysis for molecular tests as a replacement for karyotyping.

Chromosome abnormality tests tailored to the patient

- A further possible option is that women would be offered a choice and some would select a molecular test and some karyotyping. It has been assumed for this analysis that those with a high risk of fetal abnormality, around 7% of those tested, would be offered the 'new gold standard'. Offering the latter group of patients a 'new gold standard' test regime is the preferred option as this study has demonstrated that there are utility gains in terms of reduction in anxiety for all those waiting for a test result, but the group may have risks for abnormalities other than those tested for by molecular tests.
- Q-PCR is the most cost-effective test where a proportion of samples are tested using molecular methods, 7% are tested using both methods and the remainder are karyotyped. However, in laboratories with a throughput of less than 450 use of even the most cost-effective molecular test would cost 11% more than providing karyotyping for all. Above this threshold allowing patients a choice between tests is the more cost-effective option if patient preferences are similar to those in our survey population, with approximately two-thirds of patients opting for a molecular test.
- If the proportions of those choosing to be tested using molecular methods are broadly similar to our study population, 0.75–1.5 clinically significant chromosome abnormalities per 1000 women tested would remain undetected by the testing regime.

Other conclusions

- It should be recognised that one of the important potential disadvantages of molecular tests, if used to replace karyotyping within a screening programme, is that some abnormalities will remain undisclosed and women may be told that the pregnancy is normal when a rarer abnormality is present. This potential for false-negatives for a few women has to be balanced against any advantages of using molecular tests. Within a screening programme there may be serious medico-legal consequences arising from lack of understanding that the programme will miss rarer abnormalities.
- Within the option giving parental choice, protocols for case selection of high-risk patients should be developed as overall costs could rise above current UK testing costs if more than 20% of patients are offered the 'new gold standard' using FISH or 50% of patients are offered this regime using Q-PCR.
- This report has drawn conclusions based on the current cost of Q-PCR. In 2006 the cost of this test will fall by an average of 11%, when royalties are no longer payable.

Recommendations

- Within a national screening programme for Down syndrome the benefits of using a molecular test configured to detect other trisomies and numerical sex chromosome abnormalities must be an issue for debate.
- The most cost-effective testing regime, parental choice of test, is the most challenging for those counselling women before testing. Additional training for those giving parents information and counselling, particularly on the issue of false-negatives, will be necessary to achieve quality targets within a national screening programme. Suitable training packages should be developed for national use to assist achievement of equity of information to all women. In addition, health commissioners were surveyed as part of this study and their responses made it clear that even for those currently funding services there were significant gaps in knowledge. As health commissioning in the UK shifts

towards primary care, there is an urgent need to provide information about new technologies for prenatal testing for chromosome abnormalities to the individuals and organisations with responsibility for commissioning these services.
- Within the option giving parental choice, protocols for case selection of high-risk patients should be developed as overall costs could rise above current UK testing costs if more than 20% of patients are offered the 'new gold standard' using FISH or 50% of patients are offered this regime using Q-PCR.
- Currently the number of women undergoing amniocentesis is limited by the balance between the risk of a fetal abnormality and the risk of causing a miscarriage. In the near future it is likely that it will be possible to isolate fetal cells from the maternal blood-stream. The volume of samples generated will severely challenge existing services, and should be anticipated.
- The assumptions of the weighted cost-effectiveness analysis should be explored as they will clarify the results of this study.
- It was not possible to assess the impact on QoL and anxiety of replacing karyotyping with molecular tests for all women or selected groups of women within this study. This should be addressed as tests are introduced into service. Alternative mechanisms for delivery of test results should also be explored, in order to optimise the advantage of faster results.
- Currently, there is a paucity of evidence of the potential impact of false-negatives on parents, social and healthcare services. If molecular tests are to replace some karyotyping tests for some patients, further research will be needed on the impact of these changes.

Acknowledgements

This study has been a collaborative effort between many individuals and institutions. In particular, the collaboration between local hospitals and laboratories and universities has been appreciated by all those concerned. There are many people to thank.

Firstly, we would like to express our gratitude to Val Davison and her staff at the West Midlands Regional Genetic Laboratory and Consultancy Services, based at Birmingham Women's Hospital. This study would not have been possible without the valuable input from Eileen Roberts and her staff in the Cytogenetics Laboratory and we are grateful to her for allowing use of her audit data for this study. Similarly, we would like to thank Dr Geoff Smith, Dr Fiona Stewart, Mrs Louise Rauch, Mrs Maureen Collins and Mr Ian Hamilton of the Northern Ireland Regional Genetics Centre and Dr Jim Dornan, Dr Ann Harper and Dr Suresh for their support in recruiting patients.

We would especially like to thank all those local NHS staff that supported the study and gave much of their valuable time to the second phase of the study. Those who we would like to thank include the West Midlands obstetricians who supported this HTA study, particularly Mr M Brown for his support and valuable advice, and also Karen Walker, Julia White, Carol Bainbridge, Lynn Terry, Meriel Brown, Kath Parke, Sylvia Wolerston and Angela Whittaker.

We are also grateful to the NHS R&D HTA Programme for commissioning this study and to the referees for their comments. The views expressed in this report are those of the authors, who are responsible for any errors.

References

1. Tabor A, Madsen M, Obel E, Philip J, Bank J, Norgaard-Pedersen B. Randomised controlled trial of genetic amniocentesis in 4606 low-risk women. *Lancet* 1986;**i**:1287–93.
2. Mosby's dictionary of medicine, nursing and allied health. St Louis, MI: The CV Mosby Company; 1990.
3. Divane A, Carter NP, Spathas DH, Ferguson-Smith MA. Rapid prenatal diagnosis aneuploidy from uncultured amniotic fluid cells using five-colour fluorescence in situ hybridization. *Prenat Diagn* 1994;**14**:1061–9.
4. Whiteman DAH, Klinger K. Efficiency of rapid in situ hybridization methods for prenatal diagnosis of chromosomal abnormalities causing birth defects. *Am J Hum Genet* 1991;**49**(Suppl): A1279.
5. Mueller RF, Young ID. Emery's elements of medical genetics. New York and Edinburgh: Churchill Livingstone; 1995.
6. Connor JM, Ferguson-Smith MA. Essential medical genetics. Oxford: Blackwell; 1993.
7. Chitty L. Prenatal screening for chromosome abnormalities. *Br Med Bull* 1998;**54**:839–56.
8. Wald N, Kennard A, Hackshaw A, McGuire A. Antenatal screening for Down's syndrome. *Health Technol Assess* 1998;**2**(1).
9. Wald NJ, Huttly W, Hennessy C. Down's syndrome screening in the UK in 1998. *Lancet* 1999;**354**:1264.
10. Department of Health. Report of Genetics Research Advisory Group. London: Department of Health; 1995.
11. Smith N, Hau C. A six year study of the antenatal detection of fetal abnormality in six Scottish health boards. *Br J Obstet Gynaecol* 1999;**106**:206–12.
12. Cuckle H, Sehmi I. Calculating correct Down's syndrome risks. *Br J Obstet Gynaecol* 1999;**106**:371–2.
13. Advisory Committee on Genetic Testing. Prenatal Genetic Testing: Report for Consultation. 1-2-2000. London: Human Genetics Commission, Health Departments of United Kingdom; 2000.
14. Report of the RCOG Working Party on Biochemical Markers and the Detection of Down's Syndrome. London: RCOG Press; 1993.
15. Wald NJ, Kennard A, Hackshaw AK. First trimester serum screening for Down's syndrome. *Prenat Diagn* 1995;**15**:1227–40.
16. Wald N. The meaning of 'screening' [letter]. *Lancet* 1990;**336**:1587.
17. Wald NJ, Cuckle HS, Densem JW, Nanchahal K, Royston P, Chard T, *et al.* Maternal serum screening for Down's syndrome in early pregnancy. *BMJ* 1988;**297**:883–7.
18. Canick JA, Palomaki GE, Osthanondh R. Prenatal screening for trisomy 18 in the second trimester. *Prenat Diagn* 1990;**10**:546–8.
19. Howe D, Gornal R, Wellesley D, Boyle T, Barber J. Six year survey of screening for Down's syndrome by maternal age and mid-trimester ultrasound scan. *BMJ* 2000;**320**:606–10.
20. Wald NJ, Kennard A, Hackshaw A, McGuire A. Antenatal screening for Down's syndrome. *J Med Screen* 1997;**4**:181–246.
21. Huderer-Duric K, Skrablin S, Kuvacic I, Sonicki Z, Rubala D, Suchanek E. The triple-marker test in predicting fetal aneuploidy: a compromise between sensitivity and specificity. *Eur J Obstet, Gynecol Reprod Biol* 2000;**88**:49–55.
22. Kellner L, Canick J. First trimester screening for aneuploidy: serum biochemical markers. *Semin Perinatol* 1999;**23**:359–68.
23. Nicolaides KH, Snijders RJ, Cuckle HS. Correct estimation of parameters for ultrasound nuchal translucency screening [letter]. *Prenat Diagn* 1998;**18**:519–23.
24. Sjogren B. Psychological indications for prenatal diagnosis. *Prenat Diagn* 1996;**16**:449–59.
25. MRC Working Party on Evaluation of Chorionic Villus Sampling. MRC Working Party European Trial of Chorionic Villus Sampling. *Lancet* 1991;**337**:1491–9.
26. Lippman A, Tomkins D, Shime J, Hamerton J. Canadian multi-centre randomised trial of chorionic villus sampling and amniocentesis. *Prenat Diagn* 1992;**12**:385–476.
27. NEQAS. UK National External Quality Assessment Scheme: Report for Clinical Cytogenetics. 1999. URL: http://www.ukneqas.org.uk
28. Winsor EJ, Silver MP, Theve R, Wright M, Ward BE. Maternal cell contamination in uncultured amniotic fluid. *Prenat Diagn* 1996;**16**:49–54.
29. Roberts E, Ely A, Hulten M, Davison EV. The impact of FISH/PCR technology on routine prenatal diagnosis for chromosome abnormality. *Cytogenet Cell Genet* 1999;**85**:57–8.

30. Ferguson-Smith MA, Yates J. Maternal age specific rates for chromosome aberrations and factors influencing them: report of a collaborative European study on 52,965 amniocenteses. *Prenat Diagn* 1984;**4**:5–44.

31. Hook E, Gross P. Rates of mutant and inherited structural cytogenetic abnormalities detected at amniocentesis; results of 63,000 fetuses. *Ann Hum Genet* 1987;**51**:27–55.

32. Peakman DC, Moreton MF, Corn BJ, Robinson A. Chromosome mosaicism in amniotic fluid cultures. *Am J Hum Genet* 1979;**31**:149–55.

33. Chang HC, Jones OW, Masui H. Human amniotic fluid cells grown in hormone-supplemented medium: suitability for prenatal diagnosis. *Proc Natl Acad Sci USA* 1982;**79**:4796–9.

34. Chang HC, Jones OW. Reduction of sera requirements in amniotic fluid cell culture. *Prenat Diagn* 1985;**5**:305–12.

35. Held KR, Sonnichsen S. The effect of oxygen tension on colony formation and cell proliferation of amniotic fluid cells in vitro. *Prenat Diagn* 1984;**4**:171–9.

36. Hoehn H. Genetic disorders and the fetus. New York: Plenum Press; 1986.

37. Diaz VM, De La Culva P, Leal C, Alsa F. Early amniocentesis at 10–12 weeks' gestation. *Prenat Diagn* 1996;**16**:307–12.

38. Coignet JA, Van de Rijke FM, Vrolijk J, Bertheas MF, Raap AK, Tanke HJ. Automated counting of in situ hybridization dots in interphase cells in leukemia. *Leukemia* 1996;**10**:1065–71.

39. Department of Health. The genetics of common diseases. London: Department of Health; 1995.

40. Michie S, Allanson A, Armstrong D, Weinman J, Bobrow M, Marteau TM. Objectives of genetic counselling: differing views of purchasers, providers and users. *J Public Health Med* 1998;**20**:404–8.

41. Fineberg HV, Bauman R, Sosman M. Computerized cranial tomography: effect on diagnostic and therapeutic plans. *JAMA* 1977;**238**:224–30.

42. Read M, Lachs M, Feinstein A. Use of methodological standards in diagnostic test research: getting better but still not good. *JAMA* 1995;**274**:645–51.

43. Guyatt G, Tugwell P, Feeny D, Haynes R, Drummond M. A framework for clinical evaluation of diagnostic technologies. *Can Med Assoc J* 1986;**134**:587–94.

44. Marteau T, Kidd J, Cook R, Johnston M, Michie S, Shaw RW, *et al.* Screening for Down's Syndrome. *BMJ* 1988;**297**:1469.

45. Pinkel D, Landegent J, Collins C, Fuscoe J, Seagraves R, Lucas J, *et al.* Fluorescence in situ hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosomes. *Proc Natl Acad Sci USA* 1988;**85**: 9138–42.

46. Ferguson-Smith MA, Zheng YL, Carter NP. Simultaneous immunophenotyping and FISH on fetal cells from maternal blood. *Ann NY Acad Sci* 1994;**731**:73–9.

47. Hulten MA, Gould CP, Goldman ASH, Waters JJ. Chromosome in situ suppression hybridization in clinical cytogenetics. *J Med Genet* 1991;**28**:577–82.

48. Spathas DH, Divane A, Maniatist GM, Ferguson-Smith ME, Ferguson-Smith MA. Prenatal detection of trisomy 21 in uncultured amniocytes by fluorescence in situ hybridization: a prospective study. *Prenat Diagn* 1994;**14**:1049–54.

49. Mayrand PE, Ziege J, Robertson J, Hoff LB, Kronick MN. *Ann Biol Clin (Paris)* 1991;**4**:224–30.

50. Lubin MB, Blashoff JD, Wang S-W, Rotter JT, Toyoda H. Precise gene dosage determination by polymerase chain reaction; theory, methodology and statistical approach. *Mol Cell Probes* 1991;**5**:307–12.

51. Mansfield ES. Diagnosis of Down syndrome and other aneuploidies using quantitative polymerase chain reaction and small tandem repeat polymorphisms. *Hum Molec Genet* 1993;**2**:43–50.

52. Pertl B, Yau SC, Sherlock J, Davies AF, Mathew CG, Adinolfi M. Rapid molecular method for prenatal detection of Down's syndrome. *Lancet* 1994;**343**:1197–8.

53. Adinolfi M, Sherlock J, Pertl B. Rapid detection of selected aneuploidies by quantitative fluorescent PCR. *Bioessays* 1995;**17**:661–4.

54. Pertl B, Weitgasser U, Kopp S, Kroisel PM, Sherlock J, Adolfini M. Rapid detection of trisomies 21 and 18 and sexing by quantitative fluorescent multiplex PCR. *Hum Genet* 1996;**98**:55–9.

55. Szczepura AK, Kankaanpää J. Assessment of health care technologies: case studies, key concepts and strategic issues. Chichester: John Wiley; 1996.

56. Institute of Medicine. Assessing medical technologies. Washington, DC: National Academy Press: 1985.

57. Leung W, Chitayat D, Seaward G, Windrim R, Ryan G, Barrett J, *et al.* Role of amniotic fluid interphase fluorescence in situ hybridisation (FISH) analysis in patient management. *Prenat Diagn* 2001;**21**:327–32.

58. Hsu L. Prenatal diagnosis of chromosome abnormalities through amniocentesis. In: Milunsky A, editor. Genetic disorders and the fetus: diagnosis, prevention and treatment. Baltimore: John Hopkins University Press; 1992. p. 155–203.

59. Smith DK, Shaw RW, Marteau T. Informed consent to undergo serum screening for Down's syndrome: the gap between policy and practice. *BMJ* 1994;**309**:776.

60. Marteau T, Slack J, Kidd J, Shaw RW. Presenting a routine screening test in antenatal care: practice observed. *Public Health* 1992;**106**:131–41.

61. Tepperberg J, Pettenati MJ, Rao PN, Lese CM, Rita D, Wyandt H, *et al.* Prenatal diagnosis using interphase fluorescence in situ hybridisation (FISH): 2 year multi-centre retrospective study and review of literature. *Prenat Diagn* 2001;**21**:293–301.

62. Al-Jader LN, Parry-Langdon N, Smith RJW. Survey of attitudes of pregnant women towards Down syndrome screening. *Prenat Diagn* 2000;**20**:23–9.

63. Clinical genetics services in 1990 and beyond. London: Royal College of Physicians; 1991.

64. Speilberger CD, Gorsuch RL, Lushene R, Vagg PR, Jacobs GA. State–trait anxiety inventory for adults. California: Mind Garden; 1983.

65. Hall S, Bobrow M, Marteau T. Psychological consequences for patients of false negative results on prenatal testing for Down's syndrome: retrospective interview study. *BMJ* 2000;**320**:407–12.

66. Petticrew M, Sowden A, Lister-Sharp D. False negative results in screening programmes: impact and implications. *Health Technol Assess* 2000;**4**(5).

67. Drummond MF, Stoddart GL, Torrance GW. Methods for economic evaluation of health care programmes. Oxford: Oxford University Press; 1987.

68. Mutton DE, Ide RG, Alberman E. Trends in prenatal screening for and diagnosis of Down's syndrome: England and Wales. *BMJ* 1998;**317**:922–3.

69. Marteau T, Bekker H. Development of short form of the state scale of the Speilberger state trait anxiety inventory. *Br J Clin Psychol* 1992;**31**:301–6.

70. Shackley P, Cairns J. Evaluating the benefits of antenatal screening: an alternative approach. *Health Policy* 1996;**36**:103–15.

71. Ryan M, Ratcliffe J, Tucker J. Using willingness to pay to value alternative models of antenatal care. *Soc Sci Med* 1997;**44**:371–80.

72. Donaldson C, Shackley P, Abdalla M, Miedzybrodzka Z. Willingness to pay for antenatal carrier screening for cystic fibrosis. *Health Econ* 1995;**4**:439–52.

73. Donaldson C, Hundley V, Mapp T. Willingness to pay: a method for measuring preferences for maternity care? *Birth* 1998;**25**:32–9.

74. Ryan M. Valuing the benefits of health care: conjoint analysis or contingent valuation. 02/93. Aberdeen: Health Economics Research Unit, University of Aberdeen; 1993.

75. Krahn M, Mahoney J, Eckman M, Trachtrenberg J, Pauker S, Detsky A. Screening for prostate cancer: a decision analysis view. *JAMA* 1994;**272**:773–80.

76. Bush J, Chen M, Patrick D. Health status index in cost-effectiveness: analysis of PKU programme. In: Berg R, editor. Health status indexes. Chicago: Hospital Research and Education Trust; 1973. p. 172–209.

77. Javitt J, Aiello L. Cost-effectiveness of detecting and treating diabetic retinopathy. *Ann Intern Med* 1996;**124**:164–9.

78. Poon LML, Leung TN, Lau TK, Lo YMD. Prenatal detection of fetal Down's Syndrome from maternal plasma. *Lancet* 2000;**356**:1819–20.

79. Hahn S, Zhong XY, Burk MR, Troeger C, Holzgreve W. Multiplex and real-time quantitative PCR on fetal DNA in maternal plasma. A comparison with fetal cells isolated from maternal blood. *Ann N Y Acad Sci* 2000;**906**:148–52.

80. Huber K, Bittner J, Worofka B, Rosen A, Hafner E, Philipp K, *et al.* Quantitative FISH analysis and in vitro suspension cultures of erythroid cells from maternal peripheral blood for the isolation of fetal cells. *Prenat Diagn* 2000;**20**:479–86.

81. Gu S, Zhang Y, Ren C, Zhang X, Shi K, Jiang H, *et al.* [Detection of fetal cells in maternal blood by fluorescence in situ hybridization] [Chinese]. *Chung-Hua i Hsueh i Chuan Hsueh Tsa Chih* 2000;**17**:270–3.

82. Granberg M, Wikland M, Nilsson L, Hamberger L. Couples' willingness to pay for IVF/ET. *Acta Obstet Gynecol Scand* 1995;**74**:199–202.

83. Desvousges WH, Smith VK, McGivney MP. A comparison of alternative approaches for estimating recreation and related benefits of water quality improvements. EPA-230-05-83-001. Washington, DC: Office of Policy Analysis, US Environmental Protection Agency; 1983.

84. Magat WA, Viscusi WK, Huber J. Paired comparison and contingent valuation approaches to morbidity risk valuation. *J Environ Econ Manage* 1988;**15**:395–411.

85. Cairns J, Shackley P, Hundley V. Decision making with respect to diagnostic testing: a method of valuing the benefits of antenatal screening. *Med Decis Making* 2000;**16**:161–8.

86. Institute of Medicine. Assessing medical technologies. Washington, DC: National Academy Press; 1985.

Appendix 1

Development and preliminary assessment of single Q-PCR test for Down syndrome

The volume of amniotic fluid available for DNA extraction was 1.5–2.5 ml. Samples were centrifuged at 13,000 ***g*** for 5 minutes, and most of the supernatant was removed.

The first step of the procedure was to extract the DNA from the amniotic fluid sample. Genomic DNA was isolated from the mass of substance at the bottom of the tube (the cell pellet) by means of a Puregene Extraction kit (Flowgen, Lichfield); 1–6 ng/µl DNA was obtained. DNA samples were coded and the analysis was undertaken without knowledge of the patient's characteristics.

The DNA was amplified using Q-PCR, which needs to be done 24 times. The temperature at which this is performed has to be varied for each marker.

Two DNA markers (D21S11 and IFNAR) were used initially. These markers were chosen because it is a known fact that there is much variation between individuals in the length of their DNA. The test is dependent on this variation. The analysis was started using the marker D21S11. Only in those cases where this marker was uninformative (because both chromosomes contained the same DNA length) was the sample re-run with the second marker, IFNAR.

Analysis was performed using a specialised instrument, an Applied Biosystems (Warrington) 373 DNA Sequencer running Genescan 672 software. The resulting graphs, showing the lengths of the amplified markers (on the *x*-axis), as well as the amount of DNA in each position (on the *y*-axis), were then examined. Their fluorescence intensities were calculated automatically, based on the areas of the peaks seen on the graphs.

Results

In total 2167 samples were tested.

Samples without bloodstaining

In all, 2139 (98.7%) of the samples of amniotic fluid were clear, without any macroscopically detectable blood cell contamination. With the two markers D21S11 and IFNAR, 2083 (97.4%) of these samples gave informative results. For 2053 cases, two peaks of roughly equal intensity (ratio about 1) were seen (*Figure 26*). The fetus was therefore predicted to be normal, which was confirmed by karyotyping. Thirty samples were diagnosed as trisomy 21 (Down syndrome) because they showed either three peaks of equal intensity (*Figure 27*) or two peaks with a ratio of about 2 (*Figure 28*). These results were also confirmed by karyotyping.

The remaining 56 clear samples were uninformative for both markers, because they were homozygous with parental markers of the same size (*Figure 29*). Subsequent karyotyping showed that three of the fetuses had trisomy 21 and the remaining 53 were normal. For 41 of these 56 cases (39 normal, two Down syndrome) sufficient DNA remained for further analysis. These samples were re-coded and re-analysed with a third marker (D21S1270). Thirty of the 39 cytogenetically normal cases showed two Q-PCR peaks of equal intensity and were thus identified correctly as normal. The two cases of trisomy 21 identified by karyotyping were also informative with this DNA marker. They were thus diagnosed correctly by the Q-PCR method. The remaining nine (0.4%) normal cases still had uninformative Q-PCR results. However, there was not enough DNA left to be able to re-analyse them, using an additional marker. There were no false-positive or false-negative results in themselves.

Samples with bloodstaining

Twenty-eight (1.3%) of the samples of amniotic fluid showed macroscopic bloodstaining. In these cases a characteristic but inconclusive result was obtained by Q-PCR analysis with the two markers (D21S11 and IFNAR). Multiple peaks of unequal size were seen (*Figure 30*). The result obtained by Q-PCR amplification was thus unreliable and could

be misleading. With one bloodstained sample, the result, obtained by amplification of a single marker, showed two peaks with a ratio of about 1.00, an apparently normal result, but karyotyping showed the fetus to have trisomy 21. Furthermore, in another bloodstained sample, the result obtained by amplification of a single marker showed two peaks with a ratio of 2.00, which suggested a fetus with Down syndrome, but the fetus was normal on karyotyping. Analysis with a second marker, however, gave the typical pattern of bloodstained samples in both cases.

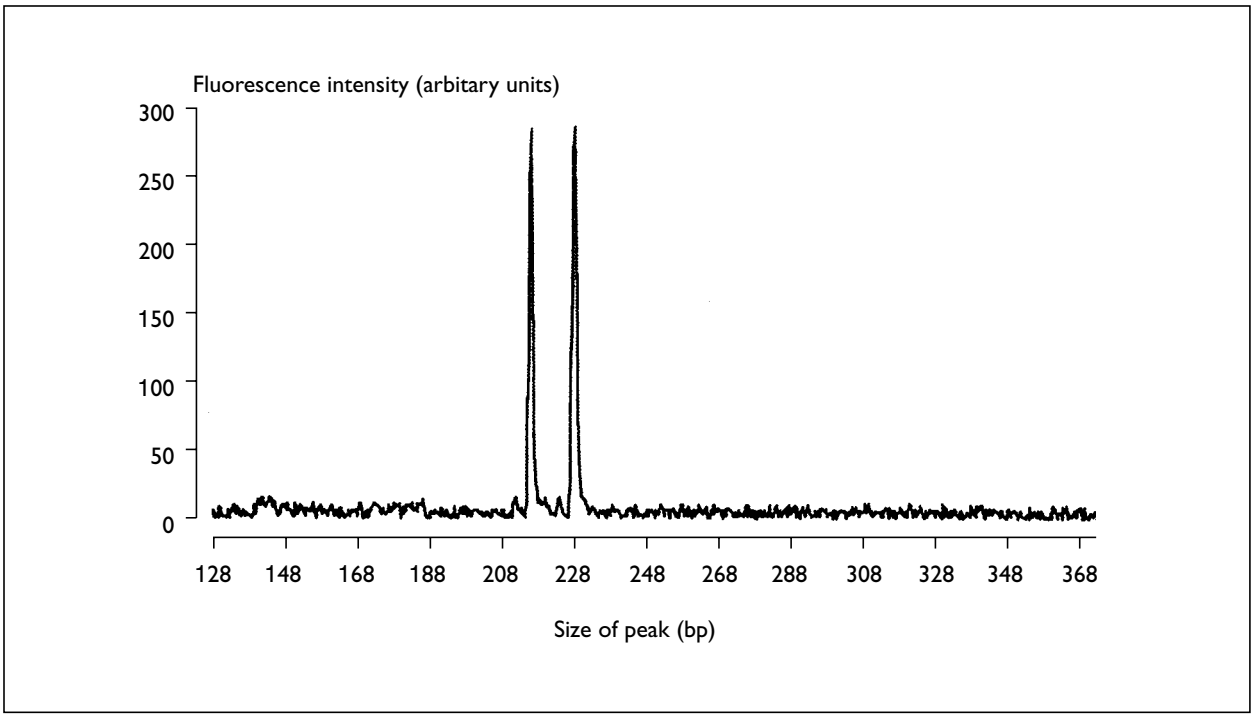


FIGURE 26 *Normal fetus. Reprinted with permission from Elsevier Science (*Lancet *1998;**352**:9–12)*

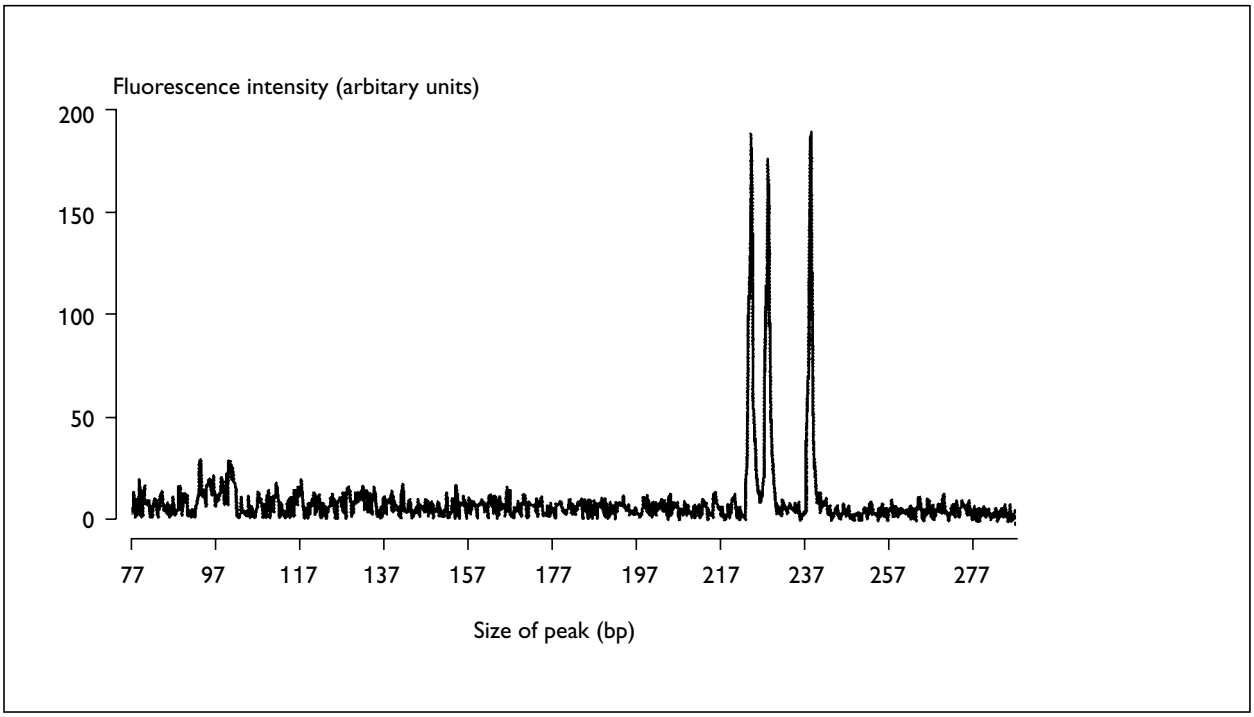


FIGURE 27 *Fetus with Down syndrome. Reprinted with permission from Elsevier Science (*Lancet *1998;**352**:9–12)*

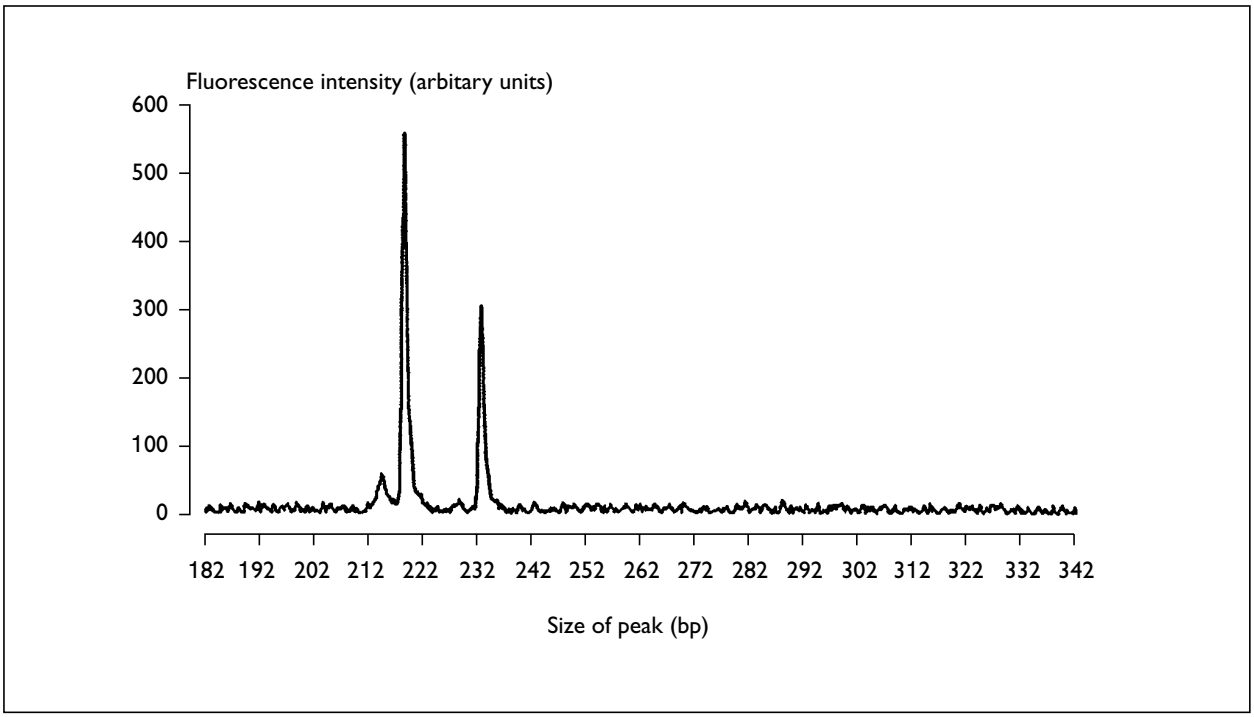


FIGURE 28 *Fetus with Down syndrome. Reprinted with permission from Elsevier Science (*Lancet *1998;**352**:9–12)*

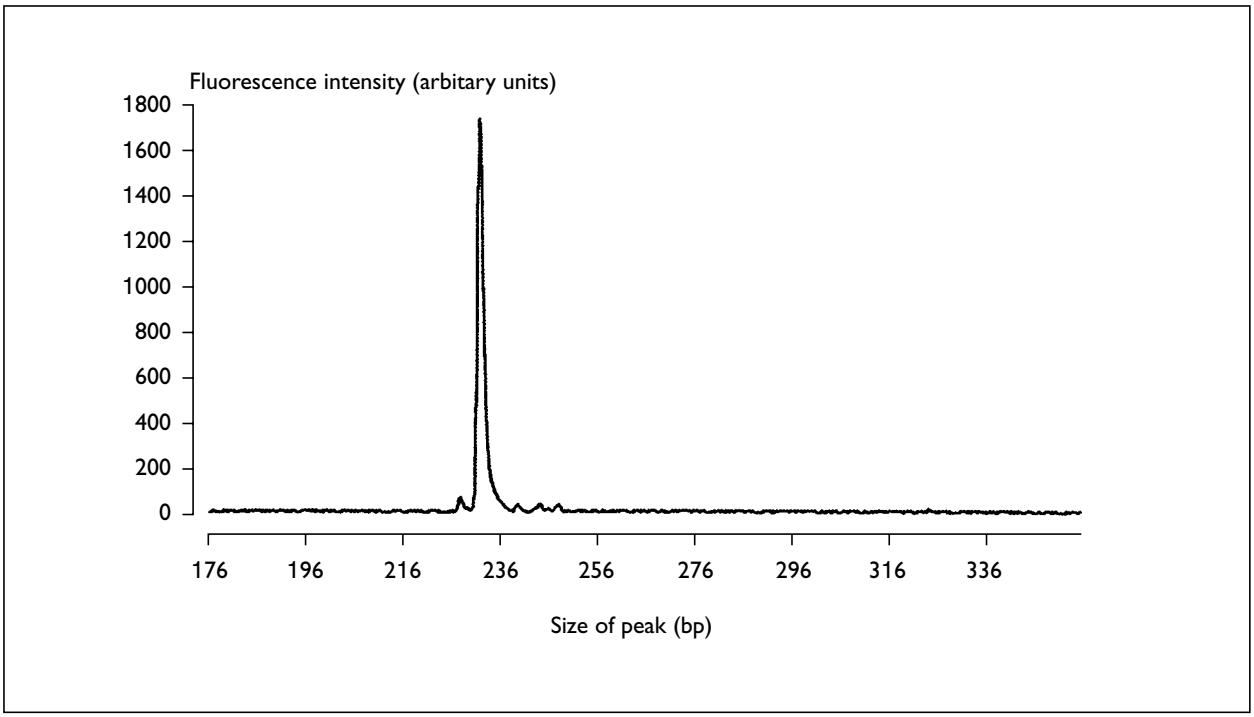


FIGURE 29 *Uninformative sample. Reprinted with permission from Elsevier Science (*Lancet *1998;**352**:9–12)*

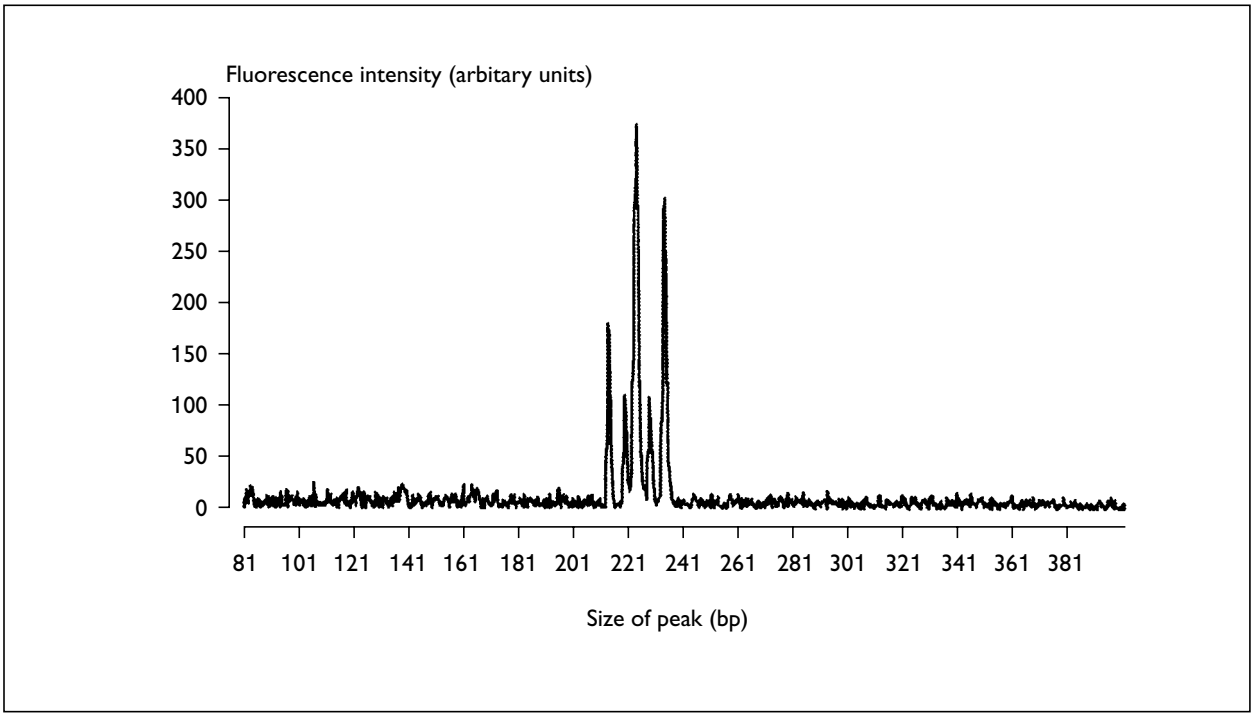


FIGURE 30 *Bloodstained amniotic fluid. Reprinted with permission from Elsevier Science (*Lancet *1998;**352**:9--12)*

Appendix 2

The number and type of abnormalities undetected if using 5-probe FISH or multiplex Q-PCR

These abnormalities would have been missed in Northern Ireland and the West Midlands if 5-probe FISH or multiplex Q-PCR was used for all women tested during 1997 and 1998.

Clinically significant chromosome abnormalities

West Midlands	Northern Ireland
46,XX,der(5)t(5;8)(p13.3;q24.1)mat	46,XY,trp(9) (pter->p13::p13->p24::p24->qter)
94,XXXX,+22,+22	mos 47,XX,+22[42]/46,XX[8]
46,XX,invdup(8)(pter;;p11p23;;qter)de novo	46,XX,del(13)(pter->q31:)
46,XX,del(5)(q?33q?34)de novo	46,XY,dup(4)(pter->q33::q32->qter)
47,XY,+mar de novo.ish der(19)(wcp19+,D1/5/19Z1+)/46,XY	
47,XY,+mar de novo.ish der(22)(D14/22Z1+)/46,XY	
47,XX,+dic(15)(q11.2)de novo	
45,X,t(1;2)(q12;p25)de novo	
46,XY,del(2)(q12.2q13)de novo	
46,XY,del(15)(q26.1)de novo	
46,XY,ins(11;21)(q23;q21q22.3)t(11;21)(q22.3;q25)de novo	
47,XX,+mar.ish der(22)(D22Z1+,wcp22+)	
46,XY,add(4q),t(8;17)(p21;q21)de novo	
46,XY,add(10q) de novo	
Autosomal mosaics	
47,XX,+20/46,XX*	
47,XY,+8/46,XY*	
47,XX,+20/46,XX*	
47,XX,+8/46,XX*	
47,XY,+der(8) (qter->p21:) de novo/46,XY*	
47,XX,+dic(15) q11.2)de novo /46,XX*	
47,XY,inv(3)(p11.2q23)pat,+8/46,XY,inv(3)(p11.2q23)pat*	
47,XY,+16/46,XY*	
47,XY,+mar/46,XY*	
47,XX,+8/46,XX*	
Sex chromosomal mosaics	
45,X/46,XX	
45,X/46,XX	
45,X[2]/46,XX[36]	
45,X[13]/46,XY[12]	
mos 47,XXY/46,XY	
45,X/46,XX	

Chromosome abnormalities with no (or unknown) clinical significance

West Midlands	Northern Ireland
46,XY,t(1;19)(q11.2;q13.1)mat	46,XX,t(4;6)(p15.2;q13)mat
45,XX,der(14;15)(q10;q10)pat	46,XX,inv(18)(pter->p11.2::q12.1->p11.2::q12.1->qter)
46,XY,inv(Y)(p11.2q11.23)pat	mos 46,XY[5]/46,XY,t(8;17)(q13;q11.2)[5]
46,XY,t(6;17)(q27;q21.3)de novo	46,XYinv(13)(q31.3q32.1)
46,X,inv(X)(q13.3q27.2)mat	46,XY,t(7;22)(p10;q10)pat
46,XX,inv(1)(p13q21)pat	46,XY,t(9;10)(q33;p15)mat
46,XX,inv(10)(p11.2q21.2)mat	
46,XY,der(9)t(1;9)(q32;p22)mat	
46,XX,t(2;6)p25;q15)de novo	
46,XY,t(13;16)(q12;p13.1)de novo	
46,XY,inv(9)(q32q34.3)mat	
46,XY,t(7;8)(p13;q11.2)pat	
46,XY,t(5;11)(p10;p10)de novo	
46,XY,t(3;6)(q25;q21)pat	
46,XX,t(8;11)(q24.1;q23.3)mat	
47,XY,+mar,pat	
46,XX,t(6;20)(p25;q11.2)mat	
46,XX,inv(13)(q12.1q14.1)pat	
45,XX,der(15;21)(q10;q10)mat	
46,XX,inv(4)(p15.3q31.3)mat	
46,XX,inv(2)(p11.2q13)mat	
46,XY,t(1;5)(p36.1;q31)de novo	
46,XX,inv(5)(p13q13)mat	
46,XY,inv(9)(q13q34)mat	
46,XY,t(1;3)(p13;p21)pat	
46,XX,inv(3)(q21.1q25.3)mat	
45,XY,der(13;14)(q10;q10)mat	
46,XY,inv(9)(q32q34)pat	
46,XX,inv(3)(q21q25)mat	
45,XX,der(14;21)(q10;q10)pat	
46,XX,inv(11)(p13p15.4)mat	
45,XY,der(13;14)(q10;q10)pat	
46,XY,inv(11)(p15.3q13.3)mat	
47,XY,+mar pat/46,XY	
45,XX,der(14;21)(q10;q10)de novo/46,XX	
46,XY,der(1)t(1;5)(p22;13),der(5)inv(5)(p13q14.2)t(1;5)pat	
46,XX,t(11;22)(q23.3;q11.2)mat	
46,XX,t(7;13)(p15;q14)mat	
46,XX t(4;11)(?p15;?q21)de novo	
46,X,inv(Y)(p11.2q11.23)	
46,XX,inv(2)(q21.1q23)mat	
46,XY,inv(3)(p11.2q23)pat	
46,XY,inv(10)(p11.2q21.2)pat	
46,XY,inv(3)(p25q21)mat	
45,XY,der(13;14)(q10;q10)pat	
46,XX,inv(2)(p11.2p21)mat	

Appendix 3

First questionnaire to women

A

New Tests for Pre-Natal Diagnosis of Chromosome Abnormalities

THE UNIVERSITY OF WARWICK

HOW TO ANSWER THIS QUESTIONNAIRE

If you do not wish to take part in the study please fill in the details on this page and then return this questionnaire to the member of staff who gave it to you.

If you are happy to take part in the study please fill in all the pages of this questionnaire and give it back to the member of staff who gave it to you. In about three days time you will receive the results of the rapid test and will be asked to fill in another questionnaire at that time. Please post this second questionnaire in the reply paid envelopes.

1. How old are you? *Tick the box*

16–19 ☐ 20–24 ☐ 25–29 ☐ 30–34 ☐ 35–39 ☐ 41–44 ☐ Over 45 ☐

2. Is this your first baby? *Tick the box*

Yes ☐ No ☐ If **No** how many previous pregnancies have you had? ☐

3. What is your ethnic origin? *Tick the box*

African--Caribbean ☐ Asian ☐ White ☐ Other ☐

4. What is your occupation?

.....

(If you don't work at the moment, say what you were doing in your last job, or put down your husband's job.)

Stop here if you are not taking part in this study

In the next three pages we are asking a series of questions which will enable us to find out about how you are feeling – these will be repeated over the next few weeks. As the leaflet you have been given explained the sample taken from the fluid surrounding your baby will be sent to the laboratory. Two tests will be carried out:

◊ a new quick test, taking three days but not giving as much information as the traditional test, the test used at the moment

◊ the traditional test, which will give more information but will take about three weeks

Please turn the page

5. In this question, we want you to imagine that you can **either** have the rapid test **or** the traditional test but not both. Which one would you choose? *(Please tick only one box)*

A result in 2 to 3 days but rare problems not picked up ☐

A result in 15 to 21 days including less common abnormalities ☐

I have no preference ☐

(go to question 7 if you ticked this box)

6. So that we can judge how strongly you feel about your choice in question 5 we want you to imagine you have to pay for the test you have chosen.
How much would you be prepared to pay to get your choice?

£0 _____
£10 _____
£20 _____
£30 _____
£40 _____
£50 _____
£60 _____
£70 _____
£80 _____
£90 _____
£100 _____
£110 _____
£120 _____
£130 _____
£140 _____
£150 _____
£160 _____
£170 _____
£180 _____
£190 _____
£200 _____
More ..... _____ say how much £.....

INSTRUCTIONS

Put a tick √ on the line if you are sure you would be willing to pay the amount

Put a ring O on the line that is the **most** you would pay

Put a cross **X** on the line if you are sure you would not be willing to pay the amount

7. We would like you to tell us about your own health state today.

By placing a tick in one box in each group below, please indicate which statement best describes your own health state today. Do not tick more than one box in each group.

Mobility

I have no problems in walking about ☐
I have some problems in walking about ☐
I am confined to bed ☐

Self-Care

I have no problems with self-care ☐
I have some problems washing or dressing myself ☐
I am unable to wash or dress myself ☐

Pain/Discomfort

I have no pain or discomfort ☐
I have moderate pain or discomfort ☐
I have extreme pain or discomfort ☐

Anxiety/Depression

I am not anxious or depressed ☐
I am moderately anxious or depressed ☐
I am extremely anxious or depressed ☐

Usual Activities (e.g. work, study, housework, family or leisure activities)

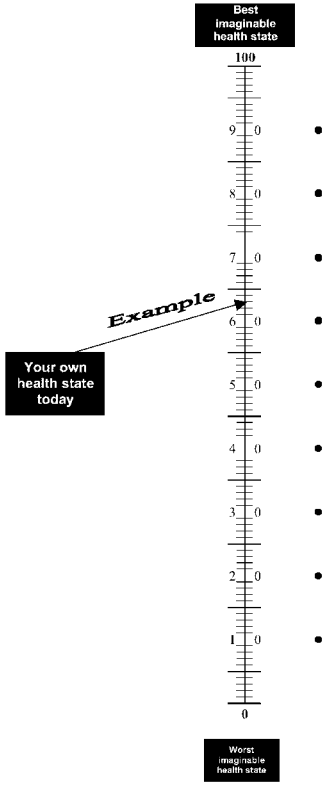
I have no problems with performing my usual activities ☐
I have some problems with performing my usual activities ☐
I am unable to perform my usual activities ☐

8. Self-Evaluation Questionnaire: How good or bad is your own health today?

To help people say how good or bad a health state is, we have drawn a scale (rather like a thermometer) on which the best state you can imagine is marked 100 and the worst state you can imagine is marked 0.

We would like you to indicate on this scale how good or bad your own health is today, in your opinion.

Please do this by drawing an arrow from the box in the middle to whichever point on the scale indicates how good or bad your health state is. An example is shown.



9. Self-Evaluation Questionnaire: Please provide the following information:

Directions: A number of statements which people have used to describe themselves are given below. Read each statement and then **circle** the appropriate number to the right of the statement to indicate (SECTION A) how you feel right now, that is, **at this moment** and (SECTION B) **how you generally feel.** There are no right or wrong answers. Do not spend too much time on any one statement.

SECTION A		NOT AT ALL	SOMEWHAT	MODERATELY SO	VERY MUCH SO
1.	I feel calm	1	2	3	4
2.	I feel secure	1	2	3	4
3.	I am tense	1	2	3	4
4.	I feel strained	1	2	3	4
5.	I feel at ease	1	2	3	4
6.	I feel upset	1	2	3	4
7.	I am presently worrying over possible misfortunes	1	2	3	4
8.	I feel satisfied	1	2	3	4
9.	I feel frightened	1	2	3	4
10.	I feel comfortable	1	2	3	4
11.	I feel self-confident	1	2	3	4
12.	I feel nervous	1	2	3	4
13.	I am jittery	1	2	3	4
14.	I feel indecisive	1	2	3	4
15.	I am relaxed	1	2	3	4
16.	I feel content	1	2	3	4
17.	I am worried	1	2	3	4
18.	I feel confused	1	2	3	4
19.	I feel steady	1	2	3	4
20.	I feel pleasant	1	2	3	4

SECTION B		ALMOST NEVER	SOMETIMES	OFTEN	ALMOST ALWAYS
21.	I feel pleasant	1	2	3	4
22.	I feel nervous and restless	1	2	3	4
23.	I feel satisfied with myself	1	2	3	4
24.	I wish I could be as happy as others seem to be	1	2	3	4
25.	I feel like a failure	1	2	3	4
26.	I feel rested	1	2	3	4
27.	I am "calm, cool, and collected"	1	2	3	4
28.	I feel that difficulties are piling up so that I cannot overcome them	1	2	3	4
29.	I worry too much over something that really doesn't matter	1	2	3	4
30.	I am happy	1	2	3	4
31.	I have disturbing thoughts	1	2	3	4
32.	I lack self-confidence	1	2	3	4
33.	I feel secure	1	2	3	4
34.	I make decisions easily	1	2	3	4
35.	I feel inadequate	1	2	3	4
36.	I am content	1	2	3	4
37.	Some unimportant thought runs through my mind and bothers me	1	2	3	4
38.	I take disappointments so keenly that I can't put them out of my mind	1	2	3	4
39.	I am a steady person	1	2	3	4
40.	I get in a state of tension or turmoil as I think over my recent concerns and interests	1	2	3	4

Appendix 4

Obstetricians' criteria for undertaking amniocentesis for specific patients

1. Abnormal 20-week ultrasound with thickened nuchal translucency and short femur length.
2. Abnormalities found in scan.
3. Bilateral large choroid plexus cysts.
4. Congenital heart – query canal defect.
5. Intrauterine growth retardation in third trimester.
6. Maternal anxiety.
7. Patient request, because of her age and previous experience in her previous pregnancy, which ended up as normal baby. (Had positive serum test in her previous pregnancy.)
8. Patient wanted all anomalies checked – will be going back to India where support for an abnormal child is apparently minimal.
9. Previous trisomy diagnosed by nuchal translucency and amniocentesis.
10. Primigravida.
11. Small for dates femur length.
12. Structural abnormality detected on ultrasound scan (fingers/heart).
13. Triple test 1:4 risk.

Appendix 5

Influence of throughput on test costs

TABLE 59 *The % additional cost (compared to cost of karyotyping) of offering karyotyping tests as add-on to molecular for proportions of population (based on throughput of UK median and quartiles)*

Test	Throughput	Proportion offered molecular test in addition to karyotyping				
		20%	40%	60%	80%	100%
FISH Down syndrome only	500	–30.18	–10.18	9.82	29.82	49.82
	1000	–33.01	–13.01	6.99	26.99	46.99
	1500	–34.00	–14.00	6.00	26.00	46.00
	5000	–35.43	–15.43	4.57	24.57	44.57
FISH 5-probe	500	5.20	23.25	45.20	65.20	85.20
	1000	3.71	23.71	43.71	63.71	83.71
	1500	3.19	23.19	43.19	63.19	83.19
	5000	2.44	22.44	42.44	62.44	82.44
Q-PCR Down syndrome only	500	–1.37	18.63	38.63	58.63	78.63
	1000	–30.76	–10.76	9.24	29.24	49.24
	1500	–41.06	–21.06	–1.06	18.94	38.94
	5000	–55.93	–35.93	–15.93	4.07	24.07
Q-PCR multiplex	500	2.17	22.17	42.17	62.17	82.17
	1000	–27.08	–7.08	12.92	32.92	52.92
	1500	–37.33	–17.33	2.67	22.67	42.67
	5000	–52.13	–32.13	–12.13	7.87	27.87

TABLE 60 *The % additional cost (compared to cost of karyotyping) of offering molecular tests as add-on to karyotyping for proportions of population based on throughput of UK median and quartiles*

Test	Throughput	Proportion offered molecular test in addition to karyotyping				
		20%	40%	60%	80%	100%
FISH Down syndrome only	500	9.96	19.93	29.89	39.85	49.82
	1000	9.40	18.80	28.19	37.59	46.99
	1500	9.20	18.40	27.60	36.80	46.00
	5000	8.91	17.83	26.74	35.65	44.57
FISH 5-probe	500	17.04	34.08	51.12	68.16	85.20
	1000	16.74	33.49	50.23	66.97	83.71
	1500	16.64	33.28	49.92	66.55	83.19
	5000	16.49	32.98	49.46	65.95	82.44
Q-PCR Down syndrome only	500	15.73	31.45	47.18	62.90	78.63
	1000	9.85	19.69	29.54	39.39	49.24
	1500	7.79	15.58	23.36	31.15	38.94
	5000	4.81	9.63	14.44	19.26	24.07
Q-PCR multiplex	500	16.43	32.87	49.30	65.74	82.17
	1000	10.58	21.17	31.75	42.33	52.92
	1500	8.53	17.07	25.60	34.13	42.67
	5000	5.57	11.15	16.72	22.29	27.87

TABLE 61 *The % additional cost (compared to cost of karyotyping) of increasing proportions offered karyotyping as a replacement for a molecular test – the 'either–or' strategy (based on throughput of UK median and quartiles)*

Test	Throughput	All molecular tests	20% m 80% k	40% m 60% k	60% m 40% k	80% m 20% k	100% Karyotyping
FISH Down syndrome only	500	–50.2	–40.1	–30.1	–20.1	–10.0	0.0
	1000	–53.0	–42.4	–31.8	–21.2	–10.6	0.0
	1500	–54.0	–43.2	–32.4	–21.6	–10.8	0.0
	5000	–55.4	–44.3	–33.3	–22.2	–11.1	0.0
FISH 5-probe	500	–14.8	–11.8	–8.9	–5.9	–3.0	0.0
	1000	–16.3	–13.0	–9.8	–6.5	–3.3	0.0
	1500	–16.8	–13.4	–10.1	–6.7	–3.4	0.0
	5000	–17.6	–14.0	–10.5	–7.0	–3.5	0.0
Q-PCR Down syndrome only	500	–21.4	–17.1	–12.8	–8.5	–4.3	0.0
	1000	–50.8	–40.6	–30.5	–20.3	–10.2	0.0
	1500	–61.1	–48.9	–36.6	–24.4	–12.2	0.0
	5000	–75.9	–60.7	–45.6	–30.4	–15.2	0.0
Q-PCR multiplex	500	–17.8	–14.3	–10.7	–7.1	–3.6	0.0
	1000	–47.1	–37.7	–28.2	–18.8	–9.4	0.0
	1500	–57.3	–45.9	–34.4	–22.9	–11.5	0.0
	5000	–72.1	–57.7	–43.3	–28.9	–14.4	0.0

k, karyotyping; m, molecular test

Appendix 6

Calculation of weighted cost-effectiveness analysis

Weighted cost per case detected is calculated in four stages:

1. Uncertainty is defined using the relative sensitivity calculated in *Table 12*. This figure gives the sensitivity of a specific test to detect all the abnormalities present including those the test is **not** designed to detect. Thus, molecular tests have a higher uncertainty than karyotyping and Q-PCR has a slightly higher uncertainty than FISH, reflecting the experimental results from stage 1 of this study, the technical evaluation phase.

   Thus,

   $\text{uncertainty}_i = 1 - \text{sensitivity}_i$

   where i is the sensitivity of the test for which the result is awaited.

   Sensitivity here is used to proxy for the lack of certainty of test result. Karyotyping has a sensitivity slightly less than unity, reflecting the fact that some tests fail to produce a result.

2. The days spent in the various states of uncertainty used assume that testing is started in the 15th week of pregnancy and that term is at 280 days. The values used are:

Period	Time waiting for result, days_i
Amniocentesis to molecular result	2
Amniocentesis to karyotyping result	17
Karyotyping result to end of pregnancy	158

3. Uncertainty days is the product of days spent in states of uncertainty, thus

   $$\text{total uncertainty relative to karyotyping} = \frac{\Sigma(1 - \text{sensitivity}_i) \times \text{days}_i}{\Sigma(1 - \text{sensitivity}_k) \times \text{days}_k}$$

   where k relates to karyotyping.

4. Cost per case is calculated as the product of unweighted cost per case detected and total uncertainty for each test configuration. It should be pointed out that throughout, a linear relationship has been assumed between the various days in states of uncertainty. The health status measurements show a trend indicating that this assumption may be capable of refinement, but this is outside the scope of this analysis.

Example:

For FISH Down only configuration:

Cost per case detected	Uncertainty	Days in state	Uncertainty days
£3184	1	2.00	2
	0.3522	15.00	5.283
	0.0001	158.00	0.0158

Therefore: Total days uncertainty = 7.30

Relative to karyotyping = 0.43

Weighted cost per case = £1366

Appendix 7

General public questionnaire

All the tests are free but now we are going to ask you to imagine that you have to pay out of your own pocket to get your choice. This is very much an exercise of the imagination which will enable us to see how strongly you feel about the option you chose on the previous page.

How much do you think you would be willing to pay to get your first choice?

Put a tick next to the amount that is the MOST you would pay to be able to have your first choice of test.

£1	_____
£10	_____
£20	_____
£30	_____
£40	_____
£50	_____
£60	_____
£70	_____
£80	_____
£90	_____
£100	_____
£120	_____
£140	_____
£160	_____
£180	_____
£200	_____
More	say how much more £.....

Remember these tests will always be free, this is an exercise of the imagination. If you have any comments can you put them in the space at the bottom of page 2

THANK YOU FOR YOUR HELP

PLEASE POST THE QUESTIONNAIRE IN THE FREEPOST ENVELOPE PROVIDED

If you have any queries about this questionnaire or about the subject discussed in the questionnaire please get in touch with Dr Grimshaw by telephoning 01203 528208 or by emailing g.m.grimshaw@warwick.ac.uk

Quicker results for genetic tests for pregnant women

Will you tell us what you think?

THE UNIVERSITY OF WARWICK

Centre for Health Services Studies

This research is being done for the NHS Research and Development Programme by Centre for Health Services Studies, University of Warwick, Coventry, CV4 7AL.

In many parts of the NHS doctors and others are trying to find out what services patients want from the NHS. We are a group of scientists, doctors and research experts who are looking for the best way to test for chromosome defects such as Down's Syndrome, in unborn babies.

We will be asking doctors and parents what they think in another survey but we would very much like you to tell us what you think. This will take about ten minutes, if you are willing to help.

Tests in pregnancy

Sadly, some babies are born who have physical or mental handicap. Tests can be done during pregnancy to check for chromosome defects that may lead to handicap. For most unborn babies this test is not necessary but all those mothers who need the test will be offered it.

One test that can show up mental or physical handicap is done by testing some of the fluid that surrounds the baby, usually around 14 weeks of pregnancy. There is a very small risk of miscarriage with this test and parents are warned of this. If the test shows there is chromosome abnormality some parents may want to decide whether to continue with the pregnancy.

What happens now

Until now parents have had to wait up to three weeks for the results of this test. The traditional test looks at all the baby's chromosomes and can tell if anything is abnormal.

New tests

New tests have been developed that can give results in two to three days. The new test works the other way round and looks for specific faults. Not all faults in the baby's make-up are tested for by the new tests but the abnormalities not tested for are very rare indeed. For example, only 4 in every 1000 babies tested will have one of these rarer problems and many of these have other tests which might pick up the problem.

Many parents will welcome the quicker result but some may be prepared to wait for the result of a test that looks at all the chromosomes.

All tests are free of charge

If doctors think it necessary, women would still be able to have the traditional test AND ALL TESTS THROUGH THE NHS ARE FREE and always will be.

Offering both the rapid and the traditional tests to all women will almost double the cost of testing and may not be necessary.

Will you tell us a little about yourself? *Tick the boxes*

How old are you?

o 16–19, o 20–24, o 25–29, o 30–34,
o 35–39, o 40–44, o 45–49, o 50 or over

Are you o Male? o Female?

How many children do you have?

o none o 1 o 2 o 3 o 4 o more than 4

Please give an indication of your household income per year

o Less than £10,000 o £10,000 to £20,000 o £20,000 to £30,000
o £30,000 to £40,000 o Greater than £40,00

What we would like you to do

Although doctors will always try to meet patient's needs we want you to **imagine** that only one test was available.

Which would you prefer it to be:

- the new rapid test which will pick up all common abnormalities and give results in two to three days,

or

- the traditional test which gives more information but normally requires a wait of at least two weeks for the results?

Please tick the box next to the test that would be your choice;

☐ **A result in 2 to 3 days but rare problems are not picked up**

☐ **A result in 15 to 21 days which includes information on rare abnormalities**

Appendix 8

Willingness to pay: contingent valuation and conjoint analysis

Information and agency – The cost–benefit approach is based on the assumption that the person paying (or willing to forego income) has perfect knowledge about the item or service and that there is no asymmetry of information between parties. However, a major feature of prenatal testing is the variability of both information giving by clinicians and understanding by the people being tested. There are indications that both clinicians and patients are not fully aware of the arguments relating to risk and diagnostic testing. Indeed there is much debate about these issues in the clinical literature.

In addition, valuation of preferences may be dependent on whether a service has previously been consumed. Furthermore, in the UK, we do not pay for healthcare and so it may be argued that respondents do not have the experience or ability to place relative values on items of healthcare expenditure. This aspect was explored in the study by measuring WTP both *ex ante* and *ex post* for women undergoing testing. The WTP value did not change at these two points of measurement.

Of all those surveyed, the health commissioners might have been expected to have the most complete knowledge about both risks and accuracy of diagnostic testing. However, overall the health commissioners' survey was the least informative, in that a high proportion of respondents were unable (or unwilling) to provide a valuation. It can be speculated that this group of respondents were more aware of the complexities of prenatal diagnosis, and therefore found it most difficult to articulate their preferences without more comprehensive information than was provided for the different tests.

Open-ended approach versus payment scale approach: Payment scales were chosen for all surveys in this study as a researcher could not be present to administer the survey in person. The need for a good response rate was considered in relation to the possible biases introduced by using payment scales. Extensive piloting preceded all questionnaires, but the possibility of scale bias must be considered when comparing cost–benefit valuations. The most likely effect of using payment scales was that respondents would produce a 'centred' result. There was no evidence of this happening in the study responses. Nevertheless, for respondents who do not pay for healthcare, the minimum and maximum of the scales may have given the respondent the impression of a 'high' value and a 'low' value in relationship to the test. In theory, using a payment scale approach as opposed to an open-ended approach would have no effect on an individual's response if the individual had perfect information on the product/service in question. However, as noted above, it is impossible to ascertain perfect information on prenatal tests for chromosome defects.

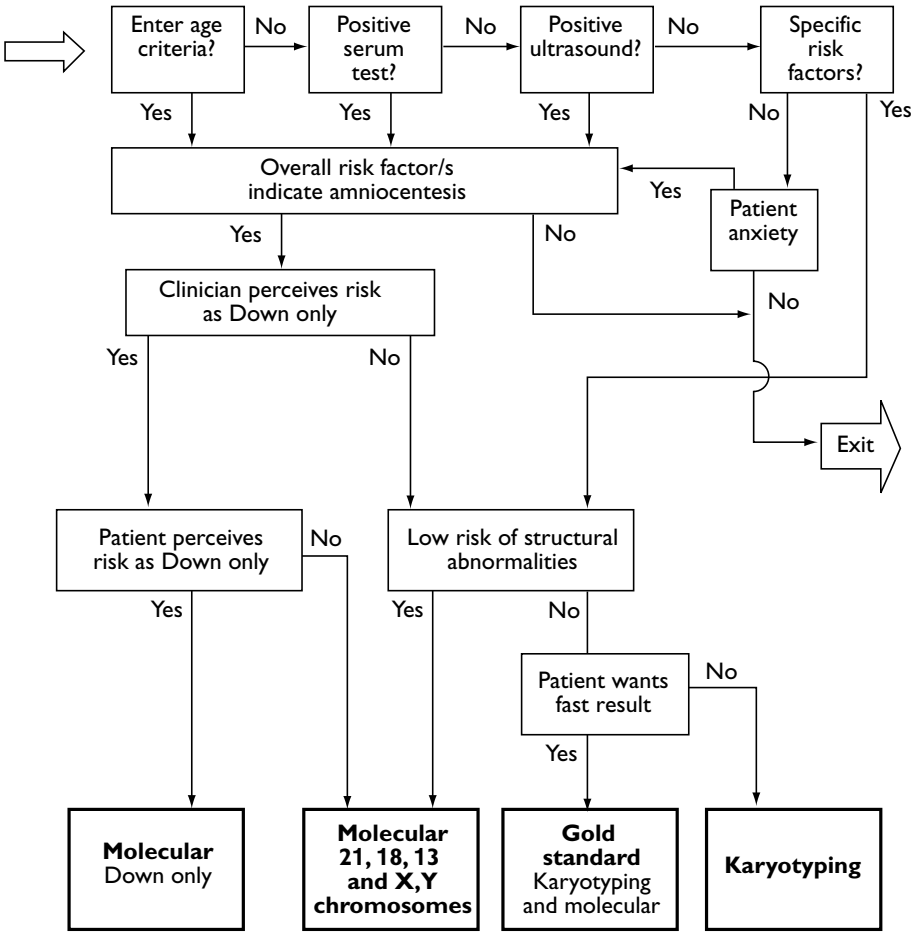
It may be argued that key stakeholders, such as the women undergoing testing and clinicians responsible for these patients, have a vested interest in the outcome of any intervention. Thus, it may be that they will present a biased valuation (their willingness to pay for the intervention) and that a more representative valuation can be provided by the general public. The general public, if asked about their willingness to pay for an intervention, are more likely to set the cost to themselves in the broader context of their overall spending patterns. Thus, the general population of Solihull was used for this survey.

The patients' surveys in particular were designed to measure health status at a traumatic and difficult time for the patient and over a stressful and anxious period of waiting. (While waiting, parents were also grappling with the decision to be taken if the test proved positive.) The patients' survey instruments had to collect demographic data, include all EuroQoL and preference questions and be administered immediately before an invasive test, in four geographically widely spaced centres within the UK. It was intended that the same method of eliciting WTP would be used with all stakeholders and thus the conditions to be met for the patients' survey dictated the choice for all surveys. In choosing a method for cost–benefit analysis, therefore, many practical as well as theoretical factors had to be balanced.

Studies similar to this one, for example evaluating antenatal cystic fibrosis screening,[72] assisted reproductive techniques,[82] and exploring competing alternatives in antenatal services,[73] have produced reliable and valid results using conjoint analysis. Where contingent analysis has been compared with contingent valuation the former gave higher values and, it has been argued, a more accurate estimation of maximum WTP.[74,83,84] However, the presentation of alternative scenarios used in some similar studies[71,85] has the disadvantage that it takes both time and concentration to read, absorb and respond to the options given. The difficulty of administering such an instrument within the context of this HTA study mitigated against the use of conjoint analysis, although perhaps theoretically it was the preferred method.

Appendix 9

Possible test strategy



Appendix 10

Aspects of assessing diagnostic technologies*

Health technology assessment is acknowledged to be more difficult for diagnostic technologies than for many other types of health technology. This is partly because diagnostic technologies are dealing with the production of information that is then mediated by an agent (e.g. a physician, surgeon or in some instances the patient) and partly because the effects of a diagnostic test or procedure on patient outcome are typically less direct than for other interventions such as pharmaceuticals. Thus, assessment of new diagnostic technologies, or extended uses of established technologies, for example developments based on improving resolution enabling imaging of smaller structures, has to determine both whether a test provides significant new diagnostic information and whether the information provided and its impact on subsequent clinical care and patient outcome offsets the costs and risks of the technology.

The impact of a diagnostic technology has therefore to be measured along a chain of inquiry such as that shown in *Table 62*. Clearly, if a diagnostic technology fails at any step in the chain then it is unlikely to be successful at a later stage. More crucially, however, success at a particular level in the hierarchy does not guarantee success at the next. Thus, an accurate test may or may not lead to more accurate diagnosis, which in its turn may or may not lead to better therapy, which may or may not result in better health for the patient, and the benefits may or may not outweigh the cost of the technology.

The immediate objective of a diagnostic technology should be to provide information about the presence, severity, and sometimes the extent, of a disease or other health condition. The diagnostic test or procedure should therefore be able to discriminate between individuals who have a particular disease or condition and those who do not, or should be capable of identifying the severity of the condition or discriminating among different extents of disease. Healthcare professionals will then be able to use this information to make decisions about the use of other interventions (including further diagnostic tests) that may in their turn affect patient health outcomes.

Many technologies used for diagnosis can also be used for population surveillance or screening. An important difference between diagnosis and screening is that, typically, diagnosis is carried out in 'symptomatic' patients who have approached the healthcare system and screening is carried out in individuals who have not sought a diagnostic test and are most likely to be asymptomatic. For a given test, used for either diagnosis or screening, the actual prevalence will have an effect on the probability that someone with a positive or negative test result has a disease or other health condition. The discussion in this appendix considers technologies used both in diagnosis and for screening and surveillance. It also focuses principally on the first two levels in the hierarchy of evaluation for a diagnostic technology shown in *Table 62,* that is determination of the technical performance of the test (assessment levels 1 and 2).

TABLE 62 *Chain of inquiry for diagnostic technologies*

Level	Assessment	Information
1	Technical capacity	Does the technology perform reliably and deliver accurate information?
2	Diagnostic accuracy	Does the technology contribute to making an accurate diagnosis?
3	Diagnostic impact	Do the results influence the pattern of subsequent use of diagnostic technologies? Does it replace other diagnostic technologies?
4	Therapeutic impact	Do the findings influence the selection and delivery of treatment?
5	Patient outcome	Does use of the diagnostic technology contribute to improved health of the patient?
6	Cost-effectiveness	Does use of the diagnostic technology improve the cost-effectiveness of healthcare compared to alternative interventions?

* Based on references 41, 55 and 86.

Precision and accuracy of diagnostic tests

A number of steps are involved in the estimation of technical performance. At the most basic level, the ability of the test to actually measure what it claims to measure must be assessed. Replicability and accuracy of test results are both important components of this element of test performance. Replicability (i.e. test precision) reflects the variance in a test result that occurs if the test is repeated on the same specimen. A highly precise test will exhibit little variance in repeat test measures, meaning that a high level of trust can be placed in a single test result in isolation. However, the fact that a test is found to be very precise does not necessarily make it acceptable as a diagnostic test. This is because a test may exhibit a high level of replicability and yet be in error. A test must therefore also be accurate; it should exhibit agreement between the test result and the true value of the biological variable being measured in the sample being tested.

As well as the precision and accuracy of a test, technical performance will also depend on the ability of the diagnostic test chosen to discriminate between patients with a particular condition or disease and those who do not have the condition or, linked with epidemiological data, to provide information on severity of disease. Most diagnostic tests measure some disease marker or surrogate (e.g. a biological factor that is associated with the disease or condition) rather than the presence or absence of the disease or condition itself. For example, in the case of hypertension, a certain agreed level of systolic and/or diastolic blood pressure is generally used as the primary diagnostic marker; hypertension is in its turn related to risk for a variety of conditions. Similarly, in the case of mild to moderate claudication, pain-free walking distance on a treadmill is widely used. In both these examples, the marker has a continuous distribution in the population tested as well as an inherent variation related to precision (*Figure 31*). The technical performance of the test will therefore depend on test variability and the appropriateness of the marker as a proxy for the condition.

For some diseases or conditions, however, it is possible to have a marker that is discrete, rather than continuous, and is, in principle, capable of discriminating populations that have, or do not have, the disease or condition, for example a microbiological test for the TB bacterium, or a prenatal test to identify a specific chromosome abnormality such as trisomy 21 (Down syndrome) in the fetus. In these cases, the technical performance of the test will be dependent on the ability of the diagnostic test used to accurately register the presence or absence of this marker.

All diagnostic tests can have four basic types of result, as shown in *Table 63*. A true-positive diagnostic test result is one that detects a marker when the disease or condition is present. A true-negative test result is one that does not detect the marker when the condition is absent. A false-positive test result is one that detects a marker when the disease or condition is absent. A false-negative test result is one that does not detect a marker when the disease or condition is present. These are usually assessed by comparing the new diagnostic technology against an appropriate gold standard (reference test).

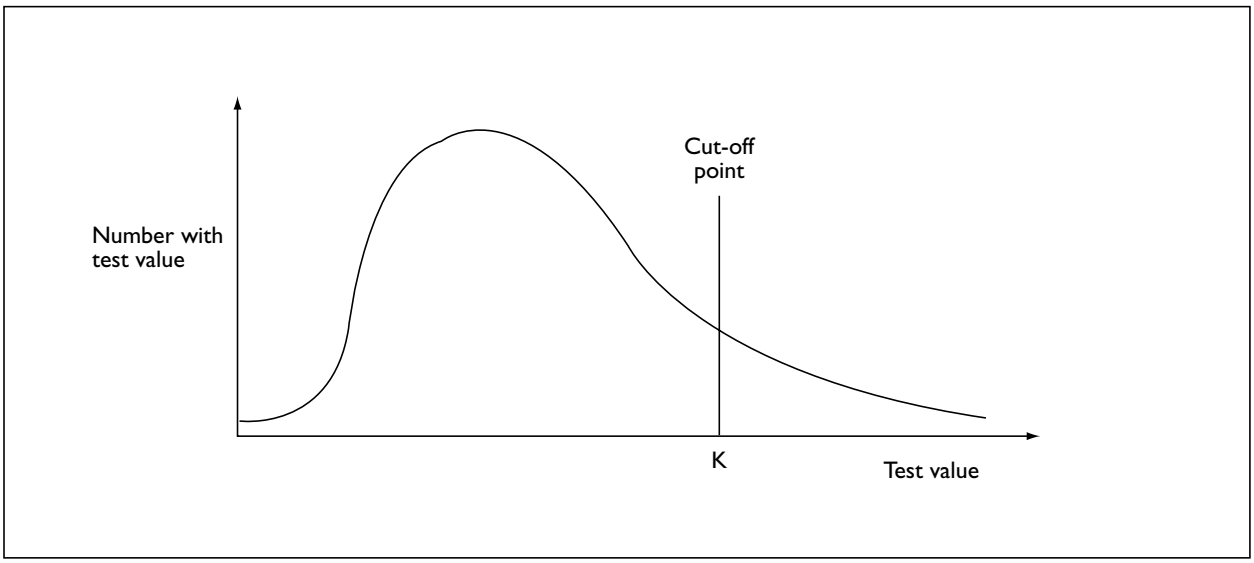


FIGURE 31 *An example of a marker with a continuous distribution in the population tested*

TABLE 63 *Possible outcomes of diagnostic tests*

Test result	Disease/condition status	
	Present	**Absent**
Positive	True-positive	False-positive
Negative	False-negative	True-negative

The relationship between a marker and the presence or absence of disease is usually not clear-cut. As *Figure 32* illustrates (for a non-discrete marker), a disease marker can be viewed as having defined distributions in populations of diseased and non-diseased people.

These distributions will arise from both variance in the populations themselves and variance in the testing procedure. The resulting distributions will commonly overlap, sometimes significantly, so that measurement of the marker will not allow complete separation of populations with and without disease.

In such a situation, clinicians and scientists will need to reach agreement about an appropriate cut-off point (K) above which disease can be assumed to be present. Clearly, no matter what cut-off point is chosen in *Figure 32* (and in *Figure 31*) it will not be possible to ensure that all patients on one side have the disease and all patients on the other do not. Instead, it can be expected that there will be some false-positive and some false-negative test results. For an established diagnostic test with a continuous marker, scientists and clinicians have usually been able to achieve a level of consensus on the appropriate cut-off level K, possibly refined for different populations (e.g. based on age, sex, etc.). Therefore, in such cases laboratories can report a test result for an individual patient as either 'positive' or 'negative'.

There will still remain an area of uncertainty, however, where the consensus cut-off has been applied as variance in the testing procedure remains. In addition, theoretically a number of different cut-off points (K) might have been selected. Moving the position of K in *Figure 32* would influence the relative probability of the different outcomes listed in *Table 63*.

Operating characteristics

The operating characteristics of a diagnostic test or procedure measure the technical performance of the technology. Operating characteristics are based on the probabilities of the four possible types of output for the diagnostic test, as shown in *Table 64*.

Sensitivity and specificity

Sensitivity represents the ability of a test to detect a disease or condition when it is present; this parameter measures the proportion of diseased

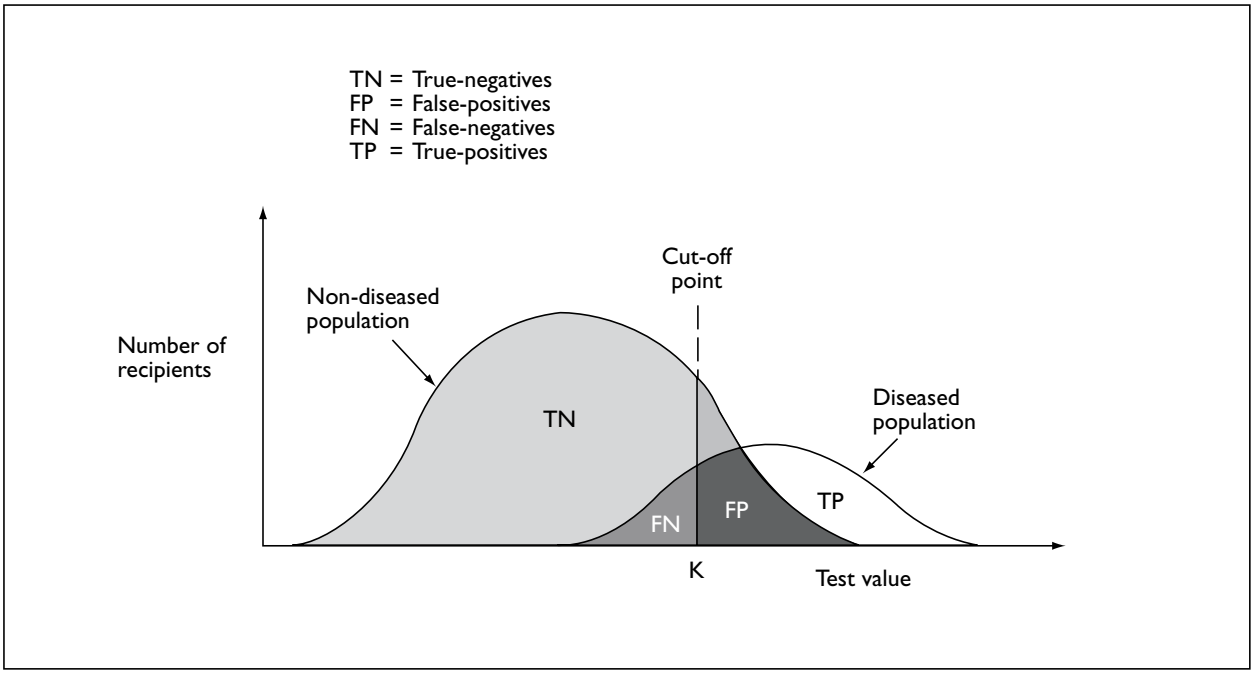


FIGURE 32 *Impact of cut-off point on test characteristics in different populations*

TABLE 64 *Operating characteristics of diagnostic tests*

Characteristic	Formula	Definition
Sensitivity	True-positives / True-positives + false-negatives	Proportion of people with condition who test positive
Specificity	True-negatives / True-negatives + false-positives	Proportion of people without condition who test negative
Positive predictive value	True-positives / True-positives + false-positives	Proportion of people with positive test who have condition
Negative predictive value	True-negatives / True-negatives + false-negatives	Proportion of people with negative test who do not have condition

patients with a positive test. Specificity represents the ability of a test correctly to exclude disease in a non-diseased person, or a condition in an individual who does not have that condition; it measures the proportion of non-diseased patients with a negative test. Sensitivity and specificity have been adopted widely because they are considered to be stable properties of a diagnostic test when derived on a broad spectrum of diseased and non-diseased patients. Under such circumstances, their values are not expected to change significantly when applied in populations with different prevalence, severity or presentation of disease.

ROC curves

A useful way of depicting these two operating characteristics for a diagnostic test is with a ROC curve, which plots the true-positive ratio (sensitivity) versus the false-positive ratio (1 – specificity) for various cut-off levels of the disease marker (i.e. values of K). *Figure 33* plots such ROC curves for hypothetical Tests A and B; the diagonal line for Test C represents a test where results are based purely on chance. Marked on the curve for Test B are estimates of sensitivity and specificity values for each of four cut-off points (K ranging from 100 to

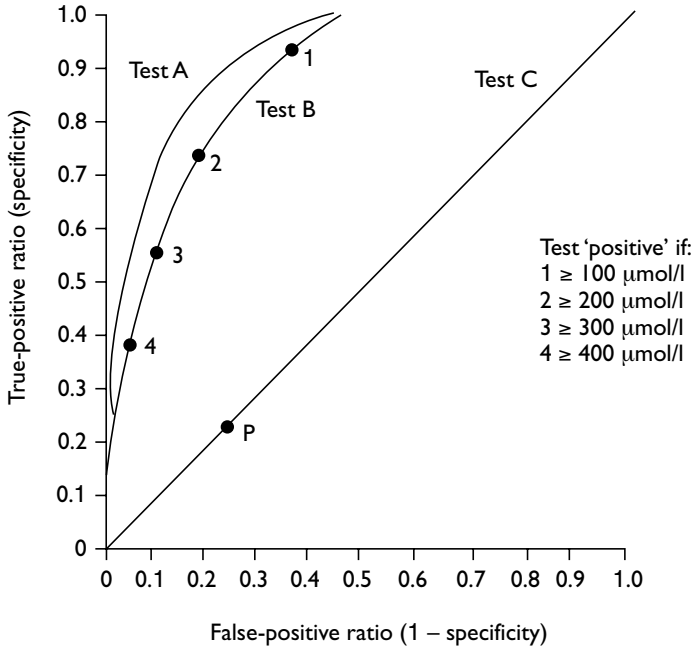


FIGURE 33 *ROC curves for three hypothetical tests*

400 µmol/l). Considering point 2, we can see that a hypothetical cut-off value of 200 µmol/l will produce a sensitivity of about 0.75 and a specificity of 0.8 (because 1 – specificity is approximately 0.2).

The curve for Test B also demonstrates that lowering the cut-off value will increase the sensitivity, but this will be at the cost of increasing the false-positive ratio (i.e. lowering the specificity). Representing the data in this way allows the specific characteristics required of a test to be considered before the final choice of cut-off is agreed by clinicians and scientists. This should include consideration of various trade-offs such as whether what is needed is high sensitivity (the ability to detect disease when it is present) or whether high specificity is required (the ability to exclude disease when it is absent). A cut-off with high specificity (to the left on the ROC curve) may be required when the aim is to confirm the presence of a disease or condition, even at the expense of a higher number of false-positives; conversely, a cut-off point with a high sensitivity may be required when screening for a disease, even though such a point is accompanied by lower test specificity.

ROC curves are also useful for comparing different tests. Because the ROC curve of Test A lies above and to the left of that for Test B in *Figure 33*, then A is shown to be a better test. This is because at every sensitivity value Test A has a lower false-positive ratio (or at every specificity value it has a higher true-positive ratio). In some instances, one test may perform better at some levels (e.g. at high sensitivity) but worse at other levels (e.g. at high specificity). In this case the ROC curves will cross and comparison is then more complicated.

Predictive values

Taken alone, sensitivity and specificity do not reveal the probability that an individual really has a disease or condition if the test is positive, or the probability that an individual does not have the condition if the test is negative. These probabilities are captured in the last two operating characteristics shown in *Table 64*. Positive predictive value is the proportion of those individuals with a positive test result who actually have the condition. Negative predictive value is the proportion of individuals with a negative test result who actually do not have the condition.

Unlike sensitivity and specificity, predictive values are not stable performance characteristics of a diagnostic test; they depend on the prevalence of the condition being examined in the population being tested. For example, as the disease prevalence (pre-test likelihood of the condition) decreases, the proportion of individuals with a positive test result who actually have the condition falls and the proportion of non-diseased patients falsely identified as having the condition rises. In contrast, as the prevalence of a disease increases, the proportion of patients with a positive test result who do in fact have the condition rises, while the proportion of patients with a negative result who do not have the disease falls. This has important implications for diagnostic tests that are used in populations with a low prevalence of disease, such as when a test is used in screening. In a situation where tests are used to screen for the presence of a disease that is sufficiently rare in the population, even tests with a high sensitivity and high specificity can have low positive predictive value, generating more false-positive than true-positive results, and therefore potentially doing as much harm as good.

Limitations in reported assessments of test performance

Currently new diagnostic technologies are usually evaluated adequately with respect to sensitivity and specificity. However, although undesirable, it is often common practice to exclude indeterminate or uninterpretable results from published evaluations of tests. Such results may occur because of technical factors (e.g. a new molecular test may be inappropriate for heavily bloodstained prenatal samples) or because the patient is not able or willing to co-operate with a diagnostic procedure (e.g. inability to repress respiration for long enough or reluctance to have eye drops for retinal photography). If these and similar cases are excluded from published results, the reported findings will overestimate the diagnostic test's actual performance once in a clinical setting (i.e. report an ideal efficacy rather than its effectiveness). Studies also rarely report other aspects of technical performance such as test replicability and repeatability.

Equally importantly, the reported performance of a new diagnostic test may not be reliable because of an element of bias in the study design. For example, the range of patients on whom a test is evaluated may be inadequate. Often a test is first assessed on patients with advanced disease and compared to young healthy controls. A test may perform well under these conditions but may not be able to discriminate patients with less advanced or less severe disease. Such selection bias will result

in inferior test performance once the test is used in a broader range of patients. Feinstein has identified four specific groups of patients in whom a test should be assessed:

- patients with a disease who are asymptomatic
- patients with symptoms representative of the spectrum of disease
- patients without the disease, but with diseases that produce similar signs and symptoms
- patients without the disease, but with diseases that affect the same organ(s) or that occur in similar anatomical locations.

In addition, a most important group are patients with the disease in whom the disease is not obvious or its presentation is atypical. Often this is the population that is least likely to be included in a test assessment. This is another cause of overstated performance being reported for a diagnostic test.

Another important limitation in assessment of test performance, particularly for tests using imaging techniques, is the fact that some element of interpretation may be part of the process of producing a test result. Thus the diagnostic performance of an ultrasound image or a diabetic retinopathy camera depends not only on the technical quality of the equipment but also on the expertise of the person interpreting the image. Reported health technology assessments, including those using the ROC curves described above, usually do not differentiate these two elements and instead address the test–interpreter combination. However, when evaluating such diagnostic technologies it may be important to separate inadequacies in the technology itself from deficiencies or difficulties inherent in interpretation, because either might be improved separately.

A further limitation when considering the performance characteristics of a test may be the lack of an appropriate gold standard (reference test). For many diseases or conditions, even the best available test (reference or gold standard) will still have some level of error, and therefore will not in fact have a sensitivity and specificity of 1.000. Evaluating a new test against an imperfect reference standard will obviously result in test performance measures that are not absolute. Ideally, any reference or gold standard should also be independent of the technology being evaluated. However, in certain circumstances the reference standard can involve expert judgement, which in its turn sometimes needs to be based, in part, on the technology being assessed. Furthermore, in some instances the reference standard used may involve subsequent confirmation at surgery or examination of tissue samples. In this situation, case selection bias can be a problem, because not all patients included in the study will necessarily have surgical or pathology reports. The evaluation results obtained may therefore not be repeatable or generalisable to the broader spectrum of patients. In some cases clinical follow-up is used as the reference standard. Clearly, such an outcome measure may be influenced by subsequent therapy.

Finally, the continuously evolving nature of many diagnostic technologies can give rise to temporal bias. The results of early assessment may be questioned in the light of subsequent improvements in the technology, or early assessments may be applied inappropriately to an improved technology.

Higher level assessment of diagnostic tests (levels 3–6)

In terms of diagnostic impact, an important issue for new diagnostic tests is the propensity for clinicians to introduce these as an add-on to existing tests. This may occur even if the new test is superior because it may be viewed as providing additional information or other benefits rather than producing replacement information. Test replacement is more likely to occur, however, where the existing test is invasive or has poor test performance.

Measurement of therapeutic impact and effect of a test on patient outcome are more difficult to assess for diagnostic technologies. Diagnostic tests are often used in combination and even a carefully constructed study design may not be able to discriminate the separate contributions of an individual test to clinical decisions and patient outcomes. Ideally, a randomised controlled trial is required in which all diagnostic pathways can be assessed, but this is often not feasible or in some cases ethical. In fact, many published assessments of diagnostic technologies are typically confined to diagnostic performance and only rarely attempt to measure clinically important impacts of diagnostic tests such as the influence on choice of therapy or the clinical outcomes following therapy. Furthermore, it is sometimes appropriate to consider the social impact of a test (e.g. screening tests) but this is not often examined.

Full economic evaluation of diagnostic tests may adopt any of the recognised forms of analysis: cost-minimisation analysis, cost-effectiveness analysis,

cost–utility analysis or cost–benefit analysis. The most common is cost-effectiveness analysis, based on measures such as the cost per case detected.

Evaluation of prenatal diagnostic tests for chromosome abnormalities

In the case of prenatal testing for chromosome abnormalities in the fetus, detection of a chromosome abnormality in fetal cells obtained through amniocentesis or CVS is used as a marker of possible physical and mental disability in the child. The current diagnostic test (karyotyping) provides a one-stage test that can identify a number of different chromosome abnormalities, visible under a microscope. Each chromosome abnormality or marker has a specific population profile in terms of the likely severity of impact on the fetus or child. However, different diagnostic markers will be able to predict the actual effect for a particular child or pregnancy to varying degrees. For example, in some cases a chromosome abnormality is known to be one that will inevitably result in a high mortality in the first few weeks of life, together with physical disabilities and severe mental disability (e.g. trisomy 13, trisomy 18). In other cases the chromosome abnormality will have a high probability of fetal loss (e.g. polyploidy where survival beyond mid-pregnancy is rare). In yet other cases, however, the abnormality, if not detected during pregnancy, may not be evident in the child unless a chromosome test is carried out later in life, for example 45,X or Turner's syndrome, which results in normal (or slightly reduced) intelligence, short stature and infertility.

For many diagnostic technologies, tests are condition-specific in that the test is designed to only identify the condition under investigation. However, other tests such as imaging procedures and broad spectrum functional biochemical tests, such as enzyme tests, may provide unexpected information over and above that requested by the clinician because of their inherent character. The diagnostic technology would, in effect, be over-diagnosing because of these inherent characteristics. Furthermore, in some cases the testing pattern is constrained by advances in the technology. For example, once automation was introduced into haematology laboratories a clinician ordering a test for suspected anaemia would receive a full blood count providing additional non-discretionary or non-requested test information (e.g. white blood cell count, platelets, etc.) as well as haemoglobin levels. The new diagnostic technology would, in effect, be over-diagnosing as a result of the introduction of automation. In both instances, the non-discretionary information need not necessarily be reported.

A similar situation currently exists for prenatal testing for chromosome abnormalities. Women who are offered a test for chromosome abnormalities in the UK (involving amniocentesis or CVS) are usually identified through a primary case selection process. This primary selection process involves assessment of risk based on parameters such as familial risk, maternal age, serum testing results and ultrasound observations. In the majority of cases the primary selection process is viewed by women as part of 'screening for Down syndrome', although the UK currently does not have a National Down Syndrome Screening Programme. However, the test carried out (karyotyping) is capable of detecting a much broader range of chromosome abnormalities.* Therefore, this also may be considered as an example of provision of non-discretionary test information or over-diagnosing, although no clinical significance may be attached to some of the abnormalities detected.

Any attempt to assess the use of molecular tests in prenatal diagnosis therefore faces an important dilemma – molecular tests are not designed to detect the full range of abnormalities that karyotyping can detect, but women and their partners are not necessarily expecting or choosing to have a test for all these other chromosome abnormalities. Therefore, a test that over-diagnoses in the terms expressed above nevertheless has to be used as the gold standard for assessment of the new molecular tests. Thus, in the present study the technical performance of a molecular test was assessed in two ways:

- In terms of its absolute technical performance – the ability of a molecular test to detect the abnormalities that the test is designed to detect compared to the gold standard (karyotyping) for these same abnormalities.

* In some cases abnormalities can be disclosed that have implications for the wider family, for example siblings of the couple tested, and issues of informed consent to testing need careful consideration before sharing of results.

- In terms of its relative technical performance – the ability of a molecular test to detect any abnormality within the range of all possible abnormalities that can be detected by the gold standard.

For higher-level assessment, beyond technical performance, the effects of prenatal diagnostic technologies on 'therapy' or on patient outcomes are typically less direct than those of other types of diagnostic technologies. The impact of a negative molecular test result (97% of tests) is principally on patient outcome, due to improved health status because of more rapid reassurance and a reduced period of anxiety, not directly on therapy. The impact of a positive test result is mediated by the parents (and to some extent clinicians and genetic counsellors) and may or may not result in termination of pregnancy.

Appendix 11

National Screening Committee's criteria for appraising the viability, effectiveness and appropriateness of a screening programme

The criteria, which are set out below, are based on the classic criteria first promulgated in a WHO report in 1966, but take into account both the more rigorous standards of evidence required to improve effectiveness and the greater concern about the adverse effects of healthcare; regrettably some people who undergo screening will suffer adverse effects without receiving benefit from the programme.

These criteria have been prepared taking into account international work on the appraisal of screening programmes, particularly that in Canada and the USA. It is recognised that not all of the criteria and questions raised in the format will be applicable to every proposed programme, but the more that are answered will obviously assist the National Screening Committee to make better evidence-based decisions.

All of the following criteria should be met before screening for a condition is initiated:

The condition

1. The condition should be an important health problem.
2. The epidemiology and natural history of the condition, including development from latent to declared disease, should be adequately understood and there should be a detectable risk factor, or disease marker and a latent period or early symptomatic stage.
3. All the cost-effective primary prevention interventions should have been implemented as far as practicable.

The test

4. There should be a simple, safe, precise and validated screening test.
5. The distribution of test values in the target population should be known and a suitable cut-off level defined and agreed.
6. The test should be acceptable to the population.
7. There should be an agreed policy on the further diagnostic investigation of individuals with a positive test result and on the choices available to those individuals.

The treatment

8. There should be an effective treatment or intervention for patients identified through early detection, with evidence of early treatment leading to better outcomes than late treatment.
9. There should be agreed evidence-based policies covering which individuals should be offered treatment and the appropriate treatment to be offered.
10. Clinical management of the condition and patient outcomes should be optimised by all healthcare providers prior to participation in a screening programme.

The screening programme

11. There must be evidence from high-quality randomised controlled trials that the screening programme is effective in reducing mortality or morbidity.
12. Where screening is aimed solely at providing information to allow the person being screened to make an 'informed choice' (e.g. Down syndrome, cystic fibrosis carrier screening), there must be evidence from high-quality trials that the test accurately measures risk. The information that is provided about the test and its outcome must be of value and readily understood by the individual being screened.
13. There should be evidence that the complete screening programme (test, diagnostic procedures, treatment/intervention) is clinically, socially and ethically acceptable to health professionals and the public.
14. The benefit from the screening programme should outweigh the physical and psychological harm (caused by the test, diagnostic procedures and treatment).

15. The opportunity cost of the screening programme (including testing, diagnosis, treatment, administration, training and quality assurance) should be economically balanced in relation to expenditure on medical care as a whole (i.e. value for money).
16. There must be a plan for managing and monitoring the screening programme and an agreed set of quality assurance standards.
17. Adequate staffing and facilities for testing, diagnosis, treatment and programme management should be made available prior to the commencement of the screening programme.
18. All other options for managing the condition should have been considered (e.g. improving treatment, providing other services), to ensure that no more cost-effective intervention could be introduced or current interventions increased within the resources available.
19. Evidence-based information, explaining the consequences of testing, investigation and treatment, should be made available to potential participants to assist them in making an informed choice.
20. Public pressure for widening the eligibility criteria for reducing the screening interval, and for increasing the sensitivity of the testing process, should be anticipated. Decisions about these parameters should be scientifically justifiable to the public.

References

Cochrane AL, Holland WW. Validation of screening procedures. *Br Med Bull* 1971;**27**:3.

Department of Health. Screening of pregnant women for hepatitis B and immunisation of babies at risk. Department of Health; 1998. (Health Service Circular: HSC 1998/127).

Gray JAM. Dimensions and definitions of screening. Milton Keynes: NHS Executive Anglia and Oxford, Research and Development Directorate; 1996.

Holland WW, Stewart S. Screening in healthcare. The Nuffield Provincial Hospitals Trust; 1990.

Sackett DL, Holland WW. Controversy in the detection of disease. *Lancet* 1975;**2**:357–9.#

Wald NJ (editor). Antenatal and neonatal screening. Oxford: Oxford University Press; 1984.

Wilson JMG, Jungner G. Principles and practice of screening for disease. Public Health Paper Number 34. Geneva: WHO; 1968.

Health Technology Assessment Programme

Prioritisation Strategy Group

Members

Chair,
Professor Kent Woods,
Director,
NHS HTA Programme, &
Professor of Therapeutics
University of Leicester

Professor Bruce Campbell,
Consultant Vascular
& General Surgeon,
Royal Devon & Exeter Hospital

Professor Shah Ebrahim,
Professor in Epidemiology
of Ageing,
University of Bristol

Dr John Reynolds,
Clinical Director,
Acute General Medicine SDU,
Oxford Radcliffe Hospital

Dr Ron Zimmern,
Director, Public Health
Genetics Unit,
Strangeways Research
Laboratories, Cambridge

HTA Commissioning Board

Members

Programme Director,
Professor Kent Woods,
Director,
NHS HTA Programme, &
Professor of Therapeutics
University of Leicester

Chair,
Professor Shah Ebrahim,
Professor in Epidemiology
of Ageing,
University of Bristol

Deputy Chair,
Professor Jon Nicholl,
Director, Medical Care
Research Unit,
University of Sheffield

Professor Douglas Altman,
Director, ICRF Medical
Statistics Group,
University of Oxford

Professor John Bond,
Director, Centre for Health
Services Research, University
of Newcastle-upon-Tyne

Professor John Brazier,
Director of Health Economics,
University of Sheffield

Dr Andrew Briggs,
Research Fellow,
Institute of Health Sciences,
University of Oxford

Ms Christine Clark,
Freelance Medical Writer,
Bury, Lancs

Professor Martin Eccles,
Professor of
Clinical Effectiveness,
University of Newcastle-
upon-Tyne

Dr Andrew Farmer,
General Practitioner &
NHS R&D Clinical Scientist,
Institute of Health Sciences,
University of Oxford

Professor Adrian Grant,
Director, Health Services
Research Unit,
University of Aberdeen

Dr Alastair Gray,
Director, Health Economics
Research Centre,
Institute of Health Sciences,
University of Oxford

Professor Mark Haggard,
Director, MRC Institute
of Hearing Research,
University of Nottingham

Professor Jenny Hewison,
Academic Unit of Psychiatry
& Behavioural Sciences,
University of Leeds

Professor Peter Jones,
University Department
of Psychiatry,
University of Cambridge

Professor Alison Kitson,
Director, Royal College of
Nursing Institute, London

Professor Sarah Lamb,
Research Professor
in Physiotherapy,
University of Coventry

Dr Donna Lamping,
Head, Health Services
Research Unit,
London School of Hygiene
& Tropical Medicine

Professor David Neal,
Department of Surgery,
University of Newcastle-
upon-Tyne

Professor Tim Peters,
Social Medicine,
University of Bristol

Professor Martin Severs,
Professor in Elderly
Health Care,
University of Portsmouth

Dr Jonathan Shapiro,
Senior Fellow, Health Services
Management Centre,
Birmingham

Dr Sarah Stewart-Brown,
Director, Health Services
Research Unit,
University of Oxford

Dr Gillian Vivian, Consultant
in Nuclear Medicine &
Radiology, Royal Cornwall
Hospitals Trust, Truro

continued

continued

Diagnostic Technologies & Screening Panel

Members

Chair,
Dr Ron Zimmern,
Director, Public Health Genetics Unit, Strangeways Research Laboratories, Cambridge

Mrs Stella Burnside,
Chief Executive,
Altnagelvin Hospitals Health & Social Services Trust,
Londonderry

Dr Paul O Collinson,
Consultant Chemical Pathologist &
Senior Lecturer,
St George's Hospital,
London

Dr Barry Cookson,
Director, Laboratory of Hospital Infection, Public Health Laboratory Service,
London

Professor Howard Cuckle,
Professor of Reproductive Epidemiology,
University of Leeds

Professor Adrian K Dixon,
Professor of Radiology,
Addenbrooke's Hospital,
Cambridge

Dr David Elliman,
Consultant in Community Child Health,
St. George's Hospital, London

Dr Tom Fahey,
Senior Lecturer in General Practice,
University of Bristol

Dr Andrew Farmer,
General Practitioner & NHS R&D Clinical Scientist,
Institute of Health Sciences,
University of Oxford

Professor Jane Franklyn,
Professor of Medicine,
University of Birmingham

Dr Antony J Franks,
Deputy Medical Director,
The Leeds Teaching Hospitals NHS Trust

Dr J A Muir Gray,
Programmes Director,
National Screening Committee,
NHS Executive, Oxford

Dr Peter Howlett,
Executive Director – Planning,
Portsmouth Hospitals NHS Trust

Dr S M Ludgate,
Medical Director,
Medical Devices Agency,
London

Professor Jennie Popay,
Professor of Sociology & Public Health,
Institute for Health Research,
University of Lancaster

Dr Susan Schonfield,
CPHM Specialist Commissioning,
Public Health Directorate,
Croydon Primary Care Trust

Mrs Kathlyn Slack,
Professional Support,
Diagnostic Imaging & Radiation Protection Team,
Department of Health,
London

Mr Tony Tester,
Chief Officer,
South Bedfordshire Community Health Council,
Luton

Dr Andrew Walker,
Senior Lecturer in Health Economics,
University of Glasgow

Professor Martin J Whittle,
Head of Division of Reproductive & Child Health,
University of Birmingham

Pharmaceuticals Panel

Members

Chair,
Dr John Reynolds,
Clinical Director,
Acute General Medicine SDU,
Oxford Radcliffe Hospital

Professor Tony Avery,
Professor of Primary Health Care,
University of Nottingham

Professor Iain T Cameron,
Professor of Obstetrics & Gynaecology,
University of Southampton

Mr Peter Cardy,
Chief Executive,
Macmillan Cancer Relief,
London

Dr Christopher Cates,
GP & Cochrane Editor,
Bushey Health Centre,
Bushey, Herts

Dr Karen A Fitzgerald,
Pharmaceutical Adviser,
Bro Taf Health Authority,
Cardiff

Dr Felicity J Gabbay,
Managing Director,
Transcrip Ltd,
Milford-on-Sea, Hants

Mr Peter Golightly,
Director, Trent Medicines Information Services,
Leicester Royal Infirmary

Dr Alastair Gray,
Director, Health Economics Research Centre,
Institute of Health Sciences,
University of Oxford

Mrs Sharon Hart,
Managing Editor, *Drug & Therapeutics Bulletin*,
London

Dr Christine Hine,
Consultant in Public Health Medicine,
Bristol South & West Primary Care Trust

Mrs Jeannette Howe,
Deputy Chief Pharmacist,
Department of Health,
London

Professor Robert Peveler,
Professor of Liaison Psychiatry,
Royal South Hants Hospital,
Southampton

Dr Frances Rotblat,
CPMP Delegate,
Medicines Control Agency,
London

Dr Eamonn Sheridan,
Consultant in Clinical Genetics, St James's University Hospital, Leeds

Mrs Katrina Simister,
New Products Manager,
National Prescribing Centre,
Liverpool

Professor Terence Stephenson,
Professor of Child Health,
University of Nottingham

Dr Richard Tiner,
Medical Director,
Association of the British Pharmaceutical Industry,
London

Professor Jenifer Wilson-Barnett, Head of Florence Nightingale School of Nursing & Midwifery, King's College,
London

Therapeutic Procedures Panel

Members

Chair,
Professor Bruce Campbell,
Consultant Vascular & General Surgeon, Royal Devon & Exeter Hospital

Professor John Bond,
Professor of Health Services Research, Centre for Health Services Research, University of Newcastle-upon-Tyne

Ms Judith Brodie,
Head of Cancer Support Service, Cancer BACUP, London

Ms Tracy Bury,
Head of Research & Development, Chartered Society of Physiotherapy, London

Mr Michael Clancy,
Consultant in A & E Medicine, Southampton General Hospital

Professor Collette Clifford,
Professor of Nursing & Head of Research, School of Health Sciences, University of Birmingham

Dr Carl E Counsell,
Senior Lecturer in Neurology, University of Aberdeen

Dr Keith Dodd,
Consultant Paediatrician, Derbyshire Children's Hospital, Derby

Mr John Dunning,
Consultant Cardiothoracic Surgeon, Papworth Hospital NHS Trust, Cambridge

Mr Jonothan Earnshaw,
Consultant Vascular Surgeon, Gloucestershire Royal Hospital, Gloucester

Professor Gene Feder,
Professor of Primary Care R&D, St Bartholomew's & the London, Queen Mary's School of Medicine & Dentistry, University of London

Professor Richard Johanson,
Consultant & Senior Lecturer, North Staffordshire Infirmary NHS Trust, Stoke-on-Trent (deceased Feb 2002)

Dr Duncan Keeley,
General Practitioner, Thame, Oxon

Dr Phillip Leech,
Principal Medical Officer for Primary Care, Department of Health, London

Mr George Levvy,
Chief Executive, Motor Neurone Disease Association, Northampton

Professor James Lindesay,
Professor of Psychiatry for the Elderly, University of Leicester

Professor Rajan Madhok,
Medical Director & Director of Public Health, North & East Yorkshire & Northern Lincolnshire Strategic Health Authority, York

Dr Mike McGovern,
Senior Medical Officer, Heart Team, Department of Health, London

Dr John C Pounsford,
Consultant Physician, Frenchay Healthcare Trust, Bristol

Professor Mark Sculpher,
Professor of Health Economics, Institute for Research in the Social Services, University of York

Dr Ken Stein,
Senior Lecturer in Public Health, Peninsular Technology Assessment Group, University of Exeter

continued

Expert Advisory Network

Members

Mr Gordon Aylward,
Chief Executive,
Association of British
Health-Care Industries,
London

Mr Shaun Brogan,
Chief Executive,
Ridgeway Primary Care Group,
Aylesbury, Bucks

Mr John A Cairns,
Reader in Health Economics,
Health Economics
Research Unit,
University of Aberdeen

Professor Nicky Cullum,
Director of Centre for
Evidence-Based Nursing,
University of York

Dr Katherine Darton,
Information Unit,
MIND – The Mental
Health Charity, London

Professor Carol Dezateux,
Professor of
Paediatric Epidemiology,
Institute of Child Health,
London

Professor Pam Enderby,
Dean of Faculty of Medicine
Institute of General Practice
& Primary Care,
University of Sheffield

Mr Leonard R Fenwick,
Chief Executive,
Freeman Hospital,
Newcastle-upon-Tyne

Professor David Field,
Professor of
Neonatal Medicine,
The Leicester Royal
Infirmary NHS Trust

Mrs Gillian Fletcher,
Antenatal Teacher &
Tutor & President,
National Childbirth
Trust, Henfield,
West Sussex

Ms Grace Gibbs,
Deputy Chief Executive
Director for Nursing,
Midwifery & Clinical
Support Services,
West Middlesex
University Hospital,
Isleworth, Middlesex

Dr Neville Goodman,
Consultant Anaesthetist,
Southmead Hospital, Bristol

Professor Robert E Hawkins,
CRC Professor & Director
of Medical Oncology,
Christie Hospital NHS Trust,
Manchester

Professor F D Richard Hobbs,
Professor of Primary Care
& General Practice,
University of Birmingham

Professor Allen Hutchinson,
Director of Public Health &
Deputy Dean of ScHARR,
University of Sheffield

Professor David Mant,
Professor of General Practice,
Institute of Health Sciences,
University of Oxford

Professor Alexander Markham,
Director,
Molecular Medicine Unit,
St James's University Hospital,
Leeds

Dr Chris McCall,
General Practitioner,
The Hadleigh Practice,
Corfe Mullen, Dorset

Professor Alistair McGuire,
Professor of Health Economics,
London School of Economics,
University of London

Dr Peter Moore,
Freelance Science Writer,
Ashtead, Surrey

Dr Andrew Mortimore,
Consultant in Public
Health Medicine,
Southampton City Primary
Care Trust

Dr Sue Moss,
Associate Director,
Cancer Screening
Evaluation Unit,
Institute of Cancer Research,
Sutton, Surrey

Mrs Julietta Patnick,
National Coordinator,
NHS Cancer
Screening Programmes,
Sheffield

Professor Chris Price,
Director of Clinical Research,
Bayer Diagnostics Europe,
Stoke Poges, Berks

Ms Marianne Rigge,
Director, College of Health,
London

Dr William Rosenberg,
Senior Lecturer &
Consultant in Medicine,
University of Southampton

Professor Ala Szczepura,
Director, Centre for
Health Services Studies,
University of Warwick

Dr Ross Taylor,
Senior Lecturer,
Department of General
Practice & Primary Care,
University of Aberdeen

Mrs Joan Webster,
Consumer member,
HTA – Expert
Advisory Network

Feedback

The HTA Programme and the authors would like to know your views about this report.

The Correspondence Page on the HTA website (http://www.ncchta.org) is a convenient way to publish your comments. If you prefer, you can send your comments to the address below, telling us whether you would like us to transfer them to the website.

We look forward to hearing from you.

Copies of this report can be obtained from:

The National Coordinating Centre for Health Technology Assessment,
Mailpoint 728, Boldrewood,
University of Southampton,
Southampton, SO16 7PX, UK.
Fax: +44 (0) 23 8059 5639 Email: hta@soton.ac.uk
http://www.ncchta.org

ISSN 1366-5278