Health Technology Assessment 2004; Vol. 8: No. 12

Clinical effectiveness and cost-effectiveness of neonatal screening for inborn errors of metabolism using tandem mass spectrometry: a systematic review

A Pandor, J Eastham, C Beverley, J Chilcott and S Paisley

March 2004

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.

Clinical effectiveness and cost-effectiveness of neonatal screening for inborn errors of metabolism using tandem mass spectrometry: a systematic review

A Pandor,[1*] J Eastham,[2] C Beverley,[1] J Chilcott[1] and S Paisley[1]

[1] The School of Health and Related Research (ScHARR), The University of Sheffield, UK
[2] Nuffield Institute for Health, University of Leeds, UK

* Corresponding author

Declared competing interests of authors: none

Published March 2004

This report should be referenced as follows:

Pandor A, Eastham J, Beverley C, Chilcott J, Paisley S. Clinical effectiveness and cost-effectiveness of neonatal screening for inborn errors of metabolism using tandem mass spectrometry: a systematic review. *Health Technol Assess* 2004;**8**(12).

Health Technology Assessment is indexed in *Index Medicus*/MEDLINE and *Excerpta Medica*/EMBASE.

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.

The research reported in this monograph was commissioned by the HTA Programme and was funded as project number 01/31/01. Technology assessment reports are completed in a limited time to inform decisions in key areas by bringing together evidence on the use of the technology concerned.

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.

ISSN 1366-5278

Published by Gray Publishing, Tunbridge Wells, Kent, on behalf of NCCHTA.
Printed on acid-free paper in the UK by St Edmundsbury Press Ltd, Bury St Edmunds, Suffolk.

Abstract

Clinical effectiveness and cost-effectiveness of neonatal screening for inborn errors of metabolism using tandem mass spectrometry: a systematic review

A Pandor,[1*] J Eastham,[2] C Beverley,[1] J Chilcott[1] and S Paisley[1]

[1] The School of Health and Related Research (ScHARR), The University of Sheffield, UK

[2] Nuffield Institute for Health, University of Leeds, UK

* Corresponding author

Objectives: To evaluate the clinical and cost-effectiveness of tandem mass spectrometry (MS)-based neonatal screening for inborn errors of metabolism (IEM).

Data sources: Fourteen electronic bibliographic databases covering biomedical, science, economic and grey literature, the reference lists of relevant articles and abstracts of conference proceedings and 18 health services research-related resources.

Review methods: This review is an update of two previous HTA reports of neonatal screening for IEM. These reports have been updated by a systematic review of published research (between 1995 and January 2002) on neonatal screening of inherited metabolic disorders using tandem MS. This was supplemented by a search for economic literature and the application of a modelling exercise to investigate the economics of using tandem MS within a neonatal screening programme in the UK.

Results: Evidence from the reviews of IEM found that the UK screening programme for phenylketonuria (PKU) was well established and there was universal agreement that neonatal screening for PKU was justified. Of the many other disorders that can be detected by tandem MS, the best candidate condition for a new screening programme was medium-chain acyl-coenzyme A dehydrogenase (MCAD) deficiency. For many other IEM that can be detected by tandem MS, robust clinical evidence was limited. Cost-effectiveness analysis using economic modelling indicated that substituting the use of tandem MS for existing technologies for the screening of PKU alone could not be justified. However, results from the economic modelling indicate that the addition of screening for MCAD deficiency as part of a neonatal screening programme for PKU using tandem MS would be economically attractive. Using an operational range of 50,000–60,000 specimens per system per year, the mean incremental cost for PKU and MCAD deficiency screening combined using tandem MS from the model was –£23,312 for each cohort of 100,000 neonates screened. This cost saving is associated with a mean incremental gain of 59 life-years. Additional economic modelling using the available evidence does not support including other inherited metabolic diseases within a neonatal screening programme at present.

Conclusions: The evidence appears to support the introduction of tandem MS into a UK neonatal screening programme for PKU and MCAD deficiency combined. Tandem MS has the potential for simultaneous multi-disease screening using a single analytical technique. Although the marginal cost of extending the programme to include other conditions may be relatively small, the application of this new technology to PKU and MCAD deficiency screening does not imply the wholesale inclusion of all disorders detectable by tandem MS. It is suggested that the primary focus of further research should be on the long-term effectiveness of treatment strategies on adverse outcomes (disabilities and impairments) under conventional management and the potential impact of early diagnosis using tandem MS.

Contents

List of abbreviations

ALD	arginosuccinase deficiency
ANOVA	analysis of variance
Arg	arginaemia
ASD	arginosuccinic acid synthetase deficiency
ASL	arginosuccinate lyase
ASS	arginosuccinate synthetase
BD	below detection limit
BKT	β-ketothiolase
CAMPA	computer-assisted metabolic profiling algorithm
CβS	cystathionine β-synthase
CCOHTA	Canadian Co-ordinating Centre for Health Technology Assessment
CCT	Current Controlled Trials
CCTR	Cochrane Controlled Trials Register
CDSR	Cochrane Database of Systematic Reviews
CEA	cost-effectiveness analysis
CEAcc	cost-effectiveness acceptability curve
CHT	congenital hypothyroidism
CI	confidence interval
CoA	coenzyme A
CPTI	carnitine palmitoyltransferase type I
CPTII	carnitine palmitoyltransferase type II
CRD	Centre for Reviews and Dissemination
CRiB	Current Research in Britain
CUA	cost–utility analysis
DARE	Database of Abstracts of Reviews of Effectiveness
DEC	Development and Evaluation Committee
DES	Development and Evaluation Services
EED	Economic Evaluation Database
EMA	ethylmalonic acidaemia
ESI	electrospray ionisation
EVPI	expected value of perfect information
F	female
FA	folic acid
FN	false negative
FP	false positive
GAD	Government Actuaries Department
GAI	glutaryl-coenzyme A dehydrogenase deficiency/glutaric aciduria type I
GAII	multiple acyl-coenzyme A dehydrogenase deficiency/glutaric aciduria type II
GCI	General Cognitive Index
GC/MS	gas chromatography/mass spectrometry
HCP	hemiplegic cerebral palsy
Hcy-Hcy	free homocysteine (the disulfide)
Hcys	homocystinuria
HDL	high-density lipoprotein
HHH	hyperornithinaemia, homocitrullinuria, hyperammonaemia syndrome
HMet	hypermethioninaemia
HMG	3-hydroxy-3-methylglutaric aciduria
HMIC	Health Management Information Consortium
HPA	hyperphenylalaninaemia
HRG	Healthcare Resource Group
HRQoL	health-related quality of life
ICER	incremental cost-effectiveness ratio
INAHTA	International Network of Agencies for Health Technology Assessment
INB	incremental net benefit
IQ	intelligence quotient
IVA	isovaleric aciduria

continued

List of abbreviations *continued*

LCAD	long-chain acyl-coenzyme A dehydrogenase
LCHAD	long-chain hydroxyacyl-coenzyme A dehydrogenase
LCPUFA	long-chain polyunsaturated fatty acid
LDL	low-density lipoprotein
M	male
MAD	multiple acyl-coenzyme A dehydrogenase
MAICER	maximum acceptable incremental cost-effectiveness ratio
2-MBCD	2-methylbutyryl-coenzyme dehydrogenase
MCAD	medium-chain acyl-coenzyme A dehydrogenase
MCC	3-methylcrotonyl-coenzyme A carboxylase
MCD	multiple carboxylase deficiency
MHP	mild hyperphenylalaninaemia
MLSO	Medical Laboratory Scientific Officer
MMA	methylmalonic aciduria
MPR	methionine-poor regimen
MRC	Medical Research Council
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MSUD	maple syrup urine disease
NA	not applicable
NCCHTA	National Co-ordinating Centre for Health Technology Assessment
NGC	National Guideline Clearinghouse
NKG	non-ketotic hyperglycaemia
NP	not performed
NR	not reported
NRR	National Research Register
NS	not significant
NSC	National Screening Committee
NTBC	2-(2-nitro-4-trifluoromethylbenzoyl)-1,3-cyclohexanedione
OHE HEED	Office of Health Economics Health Economic Evaluation Database
OLT	orthotopic liver transplantation
PCR	polymerase chain reaction
PEVPI	population expected value of perfect information
PKU	phenylketonuria
PPA	propionic aciduria
PPV	positive predictive value
PYG/PIP	pyroglutamic/pipecolic acidaemia
QALY	quality-adjusted life-year
RCT	randomised controlled trial
ReFeR	Research Findings Register
RR	relative risk
SA	succinylacetone
SCAD	short-chain acyl-coenzyme A dehydrogenase
SCHAD	short-chain hydroxyacyl-coenzyme A dehydrogenase
ScHARR	School of Health and Related Research
SD	standard deviation
SIDS	sudden infant death syndrome
SIGN	Scottish InterCollegiate Guideline Network
tfHcy	total free homocysteine
TN	true negative
TP	true positive
TRIP	Turning Research into Practice
TT1	tyrosinaemia type I
Tx	treatment
Tyr	tyrosinaemia
Vitamin B_1	thiamin
Vitamin B_{12}	cobalamin
Vitamin B_6	pyridoxine
VLCAD	very long-chain acyl-coenzyme A dehydrogenase
VOI	value of information
WHO	World Health Organization
WTE	whole-time equivalent

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

Inborn errors of metabolism are a rare group of genetic disorders that can have serious clinical consequences for an affected neonate or young infant. If undiagnosed and untreated, these disorders can cause irreversible mental retardation (ranging from mild to severe), physical disability, neurological damage and even fatality. Early detection (soon after birth) and an accurate diagnosis are very important for achieving a rapid and favourable patient outcome. Although the incidence of each specific metabolic disorder is rare, their collective importance is deemed to be of considerable public health significance.

Tandem mass spectrometry (MS) is seen as an important new technology for neonatal screening of inborn errors of metabolism. It has been demonstrated to be suitable for the reliable detection of phenylketonuria (PKU) and some other inborn errors of metabolism. This technology has the potential to screen for a range of metabolic disorders simultaneously.

The NHS R&D HTA programme commissioned two reviews of neonatal screening for inherited metabolic disorders (published in 1997). These reviews recommended further studies on the application of tandem MS to neonatal screening of inborn errors of metabolism. Both reviews were favourable to some introduction of screening for selected disorders, but with varying caveats. It was agreed that there should be no widespread introduction of the technology pending further evaluation, and the Child Health Subgroup of the National Screening Committee supported this position. Since the publication of these reports, no primary research has been undertaken in the UK. Recently, the HTA Diagnostic Technologies and Screening Panel felt that the technology was diffusing and that they needed answers more rapidly than would be obtained from proposed research. It was thought that a systematic review (to build on the two previous HTA reports) with economic modelling would be useful to bring the evidence base up to date and to identify the most urgent research needs.

Objectives

The aim of this review was to evaluate the clinical and cost-effectiveness of tandem MS-based neonatal screening for inborn errors of metabolism.

Methods

This review is an update of two previous HTA reports of neonatal screening for inborn errors of metabolism. These reports have been updated by a systematic review of recently published research (limited to studies published after 1995: the cut-off date in previous reviews) on neonatal screening of inherited metabolic diseases using tandem MS. This was supplemented by a search for economic literature and the application of a modelling exercise (based on the available evidence) to investigate the economics of using tandem MS within a neonatal screening programme in the UK.

Search strategy

New search strategies were developed based on scoping searches and strategies reported in the two previous systematic reviews. Fourteen electronic bibliographic databases covering biomedical, science, economic and grey literature were searched. The reference lists of relevant articles and abstracts of conference proceedings were checked. Eighteen health services research-related resources were also consulted via the Internet.

Inclusion and exclusion criteria

The titles, and abstracts where available, of all the articles identified by the literature searches were downloaded into a database. Any duplicates were removed. One reviewer assessed all citations and abstracts for relevance. Final decisions regarding the inclusion and exclusion criteria were based on full paper copies of manuscripts. Any uncertainties were resolved by discussion with another reviewer and/or clinical advisers.

Data extraction and quality assessment

All selected papers were read and critically appraised by a single reviewer, who extracted

relevant information from the included studies directly into an evidence table. Any uncertainties were resolved by discussion with another reviewer and/or clinical advisers. The overall strength and quality of evidence included in the review were graded according to the levels of evidence as defined in the previous systematic reviews of neonatal screening for inborn errors of metabolism.

Data synthesis

For the assessment of the clinical and cost-effectiveness evidence, formal meta-analysis techniques were not used because of the concerns over heterogeneity and poor study quality expressed in the previous systematic reviews. Instead, summary results were tabulated with detailed descriptive qualitative analyses.

Evaluation of the relative cost-effectiveness of adopting tandem MS for neonatal screening was undertaken using a probabilistic economic model. Data for the construction of the economic model were derived from previous HTA reports, other published sources and updated information on costs. A probabilistic approach was used to characterise the uncertainty within the model parameters. The model adopted an incremental net benefit approach. A value of information analysis was also undertaken to establish which elements of parameter uncertainty had the largest impact on model results.

Results

Number and quality of clinical effectiveness data

Fifteen studies were identified that evaluated the clinical effectiveness of neonatal screening for inborn errors of metabolism using tandem MS (including two abstracts and one paper that reported the same study but provided additional information or results on tandem MS-based newborn screening from different periods). Seven studies evaluated neonatal screening of amino acids and acylcarnitines using tandem MS. Four used a prospective cohort design with study durations from 2 to 3 years. Three studies did not report the study design. Of these seven studies, three were reported as abstracts and provided limited information. Four studies assessed newborn screening for medium-chain acyl-coenzyme A dehydrogenase (MCAD) deficiency using tandem MS. Three of these studies used a prospective cohort design with study durations from approximately 2 to 7 years, whereas a study from the UK, of approximately 3 years, used a retrospective cohort design. Two retrospective, analytical studies were also identified.

Assessment of clinical effectiveness

Evidence for neonatal screening of amino acids and acylcarnitines using tandem MS was primarily from observational data of large-scale prospective newborn screening programmes, from several centres outside the UK, namely, Australia, Germany and the USA. In general, newborn screening of dried blood spots for the amino acid and acylcarnitine group of disorders using tandem MS was shown to be rapid, highly sensitive (90–100%) and highly specific (99–100%). However, there was a lack of evidence regarding the false negatives and false positives for individual diagnosable disorders. The variation in the age of sampling and the heterogeneity in the choice of metabolite, as well as in thresholds used to define a positive result, limited direct comparison of the discriminative power of tandem MS between studies.

Evidence from several large-scale prospective newborn screening programmes in Australia, Germany and the USA, and retrospective data from the UK have, in general, illustrated high sensitivities and high specificities of neonatal screening for MCAD deficiency using tandem MS. However, direct comparison about diagnostic performance and outcome of tandem MS between studies was limited owing to use of various analytes and thresholds for detecting disease status and different criteria for the diagnosis of MCAD deficiency.

In most of the large-scale prospective screening studies conducted outside the UK, sampling for selected amino acids and acylcarnitines is usually performed within 72 hours after birth, whereas in the UK, blood samples are usually taken at about 6–14 days of age. The age at which screening is undertaken will affect the sensitivity and specificity of the screening process as concentrations of metabolites change over time, and this time-lapse may be detrimental for conditions that present acutely in infancy. Therefore, the evidence obtained in this review suggests that the collection of newborn blood-spots in the UK slightly earlier than the current 6–14 days may facilitate earlier detection and initiation of effective therapies. However, the earlier collection and reporting of results may influence test performance for other conditions and would underline the need for a good infrastructure for clinical follow-up, management and high-quality clinical services for identified cases and their families.

Evidence from the reviews of inborn errors of metabolism found that the UK screening programme for PKU was well established and there was universal agreement that neonatal screening for PKU was justified. The average UK incidence of PKU (classical and atypical combined) is 11.0 cases per 100,000 live births. Early dietary interventions, including dietary treatment before or during pregnancy, are effective in reducing the severity of developmental delay, and neonatal screening using tandem MS is suitable for the reliable detection of PKU.

Of the many other disorders that can be detected by tandem MS, the best candidate condition for a new screening programme was MCAD deficiency, a disorder of fatty acid metabolism (expected UK birth prevalence/incidence ranges from 4.0 to 9.9 cases per 100,000 live births). The disorder is associated with increased morbidity and mortality. Treatment for MCAD deficiency is relatively simple with dietary management, thus preventing possible early death and neurological disability. After (early) diagnosis, current management makes death rare and improves outcome. Neonatal screening data from tandem MS-based studies show that this method is robust, highly sensitive and specific for MCAD deficiency. Without screening, an unknown number may remain asymptomatic and may never experience any ill effects.

For many other inborn errors of metabolism that can be detected by tandem MS, robust clinical evidence was limited (e.g. natural history of disease, UK incidence, and uncertainties regarding the effectiveness of treatments, and sensitivity and specificity of detection using tandem MS).

Number and quality of cost-effectiveness data

Since the two previous HTA reports, no published studies on the cost-effectiveness of screening for inborn errors of metabolism using tandem MS were identified.

Cost-effectiveness analysis using economic modelling

Cost-effectiveness analysis using economic modelling indicated that substituting the use of tandem MS for existing technologies for the screening of PKU alone could not be justified. Using tandem MS would incur additional costs with no measurable increase in health benefits. However, results from the economic modelling indicate that the addition of screening for MCAD deficiency as part of a neonatal screening programme for PKU using tandem MS would be cost-effective. Using an operational range of 50,000–60,000 specimens per system per year, the mean incremental cost for PKU and MCAD deficiency screening combined using tandem MS from the model was –£23,312 (median –£14,810). This cost saving is associated with a mean incremental gain of 59 life-years for each cohort of 100,000 neonates screened.

Additional economic modelling for other conditions using probabilistic methods was undertaken, but the results are based on limited data; in particular, robust evidence on long-term outcomes, especially systematic differences in outcomes that could be attributed to screening. The evidence therefore does not support including other inherited metabolic diseases within a neonatal screening programme at present.

Conclusions

This systematic review evaluated the clinical and cost-effectiveness of neonatal screening for inborn errors of metabolism using tandem MS. The evidence appears to support the introduction of tandem MS into a UK neonatal screening programme for PKU and MCAD deficiency combined.

New technological approaches for automated processing coupled with the use of computer-assisted software would allow the analysis of hundreds of samples on a daily basis and minimise labour costs. Tandem MS has the potential for simultaneous multidisease screening using a single analytical technique. However, it is difficult to draw firm conclusions on extending the UK neonatal screening programme to all disorders detectable by tandem MS. Although the marginal cost of extending the programme to include other conditions may be relatively small, the application of this new technology to PKU and MCAD deficiency screening does not imply the wholesale inclusion of all disorders detectable by tandem MS. Robust evidence on the underlying incidence and outcomes for many of the disorders was lacking, particularly differences in long-term outcomes that could be attributed to therapies initiated as a consequence of presymptomatic detection using tandem MS.

Recommendations for research

- The economic evidence concerning the use of tandem MS for PKU and MCAD deficiency combined is very favourable. The key

assumptions underlying this analysis that may constitute areas for further research are: the future disability costs associated with symptomatic cases, and the underlying population incidence of the condition in England and Wales. However, the value of information analysis undertaken as part of the economic modelling of cost-effectiveness suggests that the overall value of obtaining this additional data is not high.

- More evidence is needed to establish the sensitivity and specificity of neonatal screening using tandem MS for other individual inborn errors of metabolism in the UK and to determine the underlying incidence of these conditions.
- Further research is needed to ascertain the natural history of some conditions and the potential economic impact of screening for other metabolic disorders. It is suggested that the primary focus of this research could be on the long-term effectiveness of treatment strategies on adverse outcomes (disabilities and impairments) under conventional management and the potential impact of early diagnosis using tandem MS.

Chapter 1

Aims of the review

Inborn errors of metabolism are a group of rare genetic disorders. In general, an inborn error of metabolism "is caused by a lack of a functional enzyme, transmembrane transporter, or similar protein, which then results in blockage of the corresponding metabolic pathway".[1]

The overall aim of this review is to evaluate the clinical and cost-effectiveness of tandem mass spectrometry (MS)-based neonatal screening for inborn errors of metabolism. More specifically the review aims:

- to update two existing systematic reviews of neonatal screening within the scope of the current review[1,2]
- to assess the viability, efficacy of test and appropriateness of screening for individual disorders detectable by tandem MS screening
- to evaluate the effectiveness and cost-effectiveness of a screening programme using tandem MS
- to identify specific research needs.

Chapter 2
Background

Background to this report

Inborn errors of metabolism are a rare group of genetic disorders that can produce serious clinical consequences for an affected neonate or young infant. If undiagnosed and untreated, these disorders can cause irreversible mental retardation (ranging from mild to severe), physical disability, neurological damage and even fatality. Early detection (soon after birth) and an accurate diagnosis are very important for achieving a rapid and favourable patient outcome. Although the incidence of each specific metabolic disorder is rare, their collective importance is deemed to be of considerable public health significance.[2]

Tandem mass spectrometry (often abbreviated as tandem MS or MS/MS) is seen as an important new technology for neonatal screening for inborn errors of metabolism. It has been demonstrated to be suitable for the reliable detection of phenylketonuria (PKU) and some other inborn errors of metabolism. This technology has the potential to screen for a range of metabolic disorders simultaneously. However, the introduction of new technologies for neonatal screening must be determined by the evidence on the need for screening each disorder (or group of related disorders) and the need for the new technology in existing screening programmes.

In the UK, the National Screening Committee (NSC) criteria for appraising the viability, effectiveness and appropriateness of a screening programme 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".[3] These appraisal criteria focus on four main areas: the condition, the test, the treatment and the screening programme, and are aimed at assisting the NSC to make better evidence-based decisions.

The NHS R&D HTA programme commissioned two reviews of neonatal screening for inherited metabolic disorders (published in 1997).[1,2] These reports recommended further studies on the application of tandem MS to neonatal screening of inborn errors of metabolism. It was agreed that there should be no widespread introduction of the technology pending further evaluation, and the Child Health Subgroup of the NSC supported this position. Both previous reports were favourable to some introduction of screening for selected conditions, but with caveats.

Pollitt and colleagues:[1]

> "There appears to be a strong case for introducing tandem MS based screening. Screening should be limited to clearly defined diseases … Given the technical complexity of the method, the large number of diseases covered, and limited experience of applying tandem MS based screening to UK populations, a 3-year pilot study is proposed …"

Seymour and colleagues:[2]

> "Screening for MCAD deficiency should be seriously considered for inclusion in newborn screening programmes. Similarly a case can be made for the introduction of screening for GAI. … Such screening is dependent on the introduction of tandem MS technology … Tandem MS could simultaneously detect other selected disorders."

Various newborn-screening programmes in Australia (New South Wales Newborn Screening Program)[4,5] and the USA (Massachusetts Newborn Screening Program,[6] New England Newborn Screening Program,[7] North Carolina Newborn Screening Program,[8] Wisconsin Newborn Screening Program)[9] have introduced tandem MS technology for the detection of various amino acid and acylcarnitine disorders.

A recent report from the Centers for Disease Control and Prevention, USA, "Using tandem mass spectrometry for metabolic disease screening among newborns", has also been published.[6] However, this focused largely on the logistical matters involved in implementing the technology, rather than on the policy questions of what it should be used for. This report covers a range of issues, including good laboratory practice and quality assurance, but it did identify some key areas of uncertainty:

- which conditions should be included in a screening programme

- uncertainties over the success of treatment for some conditions
- difficulties in cut-offs on conditions with varying degrees of severity, and lack of data on the natural history of milder variants.

There was also an agreement within the report that "long-term studies are needed to evaluate whether outcomes are improved as a result of MS/MS screening and early diagnosis".[6]

Since the publication of the two previous HTA reports and the introduction of tandem MS-based technology in various newborn-screening programmes internationally, no primary research has been undertaken in the UK. Recently, the HTA Diagnostic Technologies and Screening Panel felt that the technology was diffusing and that they needed answers more rapidly than would be obtained from proposed research. In the short term, it was suggested that the HTA commission a systematic review (building on the two previous HTA reviews and bringing them up to date) with economic modelling. This report systematically reviews the evidence of clinical effectiveness and cost-effectiveness (with economic modelling) of neonatal screening for inborn errors of metabolism using tandem MS.

Description of new intervention: tandem MS

Mass spectrometry is an analytical technique that is used to identify unknown compounds, quantify known materials, and elucidate the structural and physical properties of ions. A mass spectrometer is a device that separates and quantifies ions based on their mass-to-charge ratios. Mass spectrometers measure weight (i.e. mass) electronically and display results in the form of a mass spectrum. A mass spectrum is a graph that shows each specific molecule by weight and how much of each molecule is present.

Tandem MS is based on the use of two mass spectrometers connected in series by a chamber (known as a collision cell), which breaks down molecules into their constituent pieces. A sample is sorted and weighed in the first mass spectrometer, and fragmented within the collision cell, then the pieces are further sorted and weighed in the second mass spectrometer. The amount of sample required is minute (tandem MS instruments are very sensitive in that compounds in fractions of nanomoles can be separated and identified) and the entire procedure, from ionisation and sample injection to data acquisition by computer, is very rapid.

Different types of mass spectrometer are available. One distinguishing feature of a tandem MS system is the way in which the compound is placed into the tandem MS [e.g. fast atom bombardment, liquid secondary ion or electrospray ionisation (ESI)]. The preferred type of sample introduction is by electrospray ionisation, as this allows a continuous flow approach. Therefore, a single tandem MS instrument can process a large volume of specimens, through the use of an automated sampler. Additional improvements to the technology during the 1990s have facilitated the ease with which tandem MS can be applied. Tandem MS instruments now include the use of automated interpretation systems.

One of the perceived advantages of tandem MS for neonatal screening is that the technology allows for the simultaneous detection of a much wider range of metabolic disorders than conventional methods. Analysis for these additional conditions can be undertaken using the same blood-spot sample provided for an existing screened disorder (such as PKU); no additional specimen collection or sample preparation is required.

At present, other neonatal test methods are not available for detection of acylcarnitine disorders, such as medium-chain acyl-coenzyme A (CoA) dehydrogenase (MCAD) deficiency, and so tandem MS would be required to detect these disorders routinely as part of a dedicated screening programme. *Table 1* shows the different types of metabolic disorder that can be detected from blood-spot analysis using a tandem MS instrument.

A series of recommendations about important technical and laboratory practices in relation to tandem MS (sample preparation, instrumentation calibration, quality assurance, etc.) has recently been published by the Centers for Disease Control and Prevention in the USA.[6]

Neonatal screening for inborn errors of metabolism in the UK

The neonatal screening programme in the UK currently screens for two biochemical disorders, PKU and congenital hypothyroidism (CHT). A national screening programme for PKU has existed since 1969 and in 1981 screening for CHT was added.

TABLE 1 *Metabolic disorders detectable in newborns using tandem MS*[a]

Primary disease/screen	Diseases detected
Amino acid scan	Phenylketonuria (PKU) Maple syrup urine disease (MSUD) Argininosuccinic aciduria Tyrosinaemia type I Homocystinuria (by hypermethioninaemia) 5-Oxoprolinuria
Urea cycle scan	Ornithine carbamoyltransferase and other urea cycle defects by secondary increase in glutamine Citrullinaemia Hyperornithinaemias
Acylcarnitine scan	Medium-chain acyl-CoA dehydrogenase (MCAD) deficiency Defects of long-chain fatty acid catabolism Defects of branched-chain fatty acid acyl-CoA catabolism Glutaric acidaemia type I (GAI) Glutaric acidaemia type II and related disorders Propionic acidaemia Methylmalonic acidaemia (but possibly not mildest forms) Isovaleric acidaemia

Source: Pollitt *et al.*[1]
[a] The list is not all-inclusive but serves only as a guideline.

PKU is caused by an enzyme deficiency leading to defective hydroxylation of phenylalanine to tyrosine in the liver. The result is an accumulation of phenylalanine in the blood and tissues. Classical PKU is the severe form of the disease resulting from complete deficiency of the relevant enzyme. If untreated this will lead to severe mental disability associated with other neurological abnormalities. Patients usually appear normal in the first few months of life, but signs of delayed development will be apparent by 6–12 months. Early intervention prevents neurological damage, as the condition is treatable using dietary therapy.

CHT is caused by defective functioning of the thyroid gland from birth. Babies with this condition are likely to present clinically at birth, although the degree of severity of the condition will vary. A severely affected infant will have characteristic features, feeding difficulties, constipation and generally sluggish behaviour. However, this severe form is relatively rare and the majority of cases may not be recognised clinically during the neonatal period. If left untreated growth will be retarded and there will be mild to severe mental retardation.

Screening for PKU developed following successful attempts to control the condition by diet and the recognition that the best results of dietetic treatment were obtained if treatment was started in infancy. The earliest tests used urine samples; blood-based screening was then instigated by Robert Guthrie, who developed a bacterial inhibition assay for phenylalanine in dried blood spots collected on a filter-paper card (Guthrie card). In some areas of the UK Guthrie's assay was first used on urine dried on filter paper, but by 1969 it had become clear that a blood-based Guthrie approach was superior and this type of screening was adopted nationally. The test for CHT measures thyroid-stimulating hormone.

The screening programme is based on a single sample of capillary blood collected by heel prick from all infants between 6 and 14 days of age. The samples are collected by a midwife or health visitor as either blood spots on a filter-paper card (the most common method) or in a small heparinised capillary tube. A single blood sample is common to screening for both conditions. Small discs for testing are punched out of the dried blood spots at the screening laboratory and these are used for the assays.

In the UK PKU and CHT are the only disorders screened for universally. Evidence indicates that the screening programme is very effective, with few cases having been missed.[10] Early detection and treatment of individuals at risk of these disorders can reduce or prevent serious neurological and developmental damage.

During 1999, the Office for National Statistics estimated the number of all live births as 700,192 annually in the UK [distribution of live births: England, 589,468; Wales, 32,111; Scotland (includes birth to women normally resident outside Scotland), 55,147 and Northern Ireland (excludes births to women normally resident outside Northern Ireland), 22,957].[11] Seymour and colleagues[2] reported that in total about 250 babies are born each year with an inborn error of metabolism (including PKU) and consider this to be of considerable public health significance.

Chapter 3

Methods for reviewing evidence

Method overview

This review is essentially divided into two parts:

- systematic reviews of the clinical evidence
- a systematic review of the economic evidence and an economic modelling exercise.

Systematic reviews of the clinical evidence

The following systematic reviews were undertaken to evaluate the evidence on the clinical effectiveness of neonatal screening for inborn errors of metabolism using tandem MS.

- First, a review of the clinical evidence examined recently published research that assessed the efficacy of tandem MS as a neonatal screening technology for inborn errors of metabolism.
- Second, for a number of individual inborn errors of metabolism detectable by tandem MS technology, several systematic reviews of the recently published literature (11 separate systematic reviews covering 23 individual inborn errors of metabolism) were undertaken to establish the burden of inborn errors of metabolism (natural history, population incidence, prevalence, morbidity and mortality) and the effectiveness of interventions for these conditions (treatment and outcome). Where there was no additional epidemiological information or evidence on treatments, data from the two previous systematic reviews[1,2] were reported.

Systematic review of the economic evidence and an economic modelling exercise

The systematic review of the economic evidence assessed any published research on the economics of neonatal screening using tandem MS. This section also provided an economic model that examined the potential cost-effectiveness of replacing the existing national programme technologies (for PKU) with tandem MS and the implications of then extending screening to additional disorders. Where possible, published and routine data sources were used, supplemented where necessary with expert judgement. Using this model, a sensitivity analysis was undertaken with the objective of identifying the key uncertainties underlying the economics of screening. The aim of this was to identify whether available information and evidence was sufficient for the purpose of commissioning or whether further research would be of value.

Search strategies

Background context

This current review is an update (limited to studies published after 1995, the cut-off date in the previous reviews) of two existing systematic reviews of neonatal screening for inborn errors of metabolism.[1,2] In the Pollitt review, the authors conducted an exhaustive literature review of various electronic databases covering biomedical, economic and psychological literature between 1966 and 1995. Seymour and colleagues[2] carried out a comprehensive literature review of various computerised databases covering biomedical literature between 1966 and June 1996. Both groups also conducted manual searches of textbooks, conference proceedings and reference lists of articles.

In the present review, new search strategies were developed based on scoping searches and strategies reported in the two existing systematic reviews.[1,2] The search aimed to identify all references related to the clinical effectiveness of neonatal screening for inborn errors of metabolism using tandem MS. In addition, updated searches were conducted on each of the 23 major inborn errors of metabolism detectable by tandem MS in terms of epidemiology, diagnosis, treatment and screening. This was supplemented by a search for economic literature and an economic modelling exercise to estimate the relative cost-effectiveness using tandem MS within a neonatal screening programme in the UK (further methodological details are provided in Chapter 7). All literature searches were conducted between November 2001 and January 2002.

Review of tandem MS

Sources searched

Twelve electronic bibliographic databases were searched, covering biomedical, science, and grey

literature. A list of databases is provided in Appendix 1. In addition, the reference lists of relevant articles were checked (including abstracts of conference proceedings) and various health services research-related resources were consulted via the Internet. These included health technology assessment organisations, guideline-producing bodies, and generic research and trials registers. A list of these additional sources is given in Appendix 2.

Search terms

A combination of free-text and thesaurus terms was used. 'Neonatal screening' search terms were combined with 'inborn errors' terms and 'tandem mass spectrometry' terms. Copies of the search strategies used in the major databases are included in Appendix 3.

Search restrictions

No language, publication type or study design restrictions were applied. In addition, no date restrictions were used owing to the small number of studies identified during the scoping process (during the period from 1966 to November 2001) for neonatal screening using tandem MS.

Review of individual inborn errors of metabolism

Sources searched

Supplementary searches for each of the individual inborn errors of metabolism were conducted in the MEDLINE database. Manual searches of textbooks and reference lists of relevant articles were also examined to identify additional relevant citations.

Search terms

A combination of free-text and thesaurus terms was used for each of the 23 major inborn errors of metabolism [e.g. phenylketonuria: phenylketonurias (exploded MeSH term) or phenylketonuria$]. Copies of the search terms are provided in Appendix 4.

Search restrictions

Searches were undertaken from 1995 (scoping searches generated numerous citations over the period from 1966 to January 2002, and studies prior to 1995 would have already been identified by the existing systematic reviews: the cut-off date in the Pollitt[1] review was 1995; that in Seymour[2] review was June 1996) to the present (January 2002). Searches were further restricted to epidemiology, diagnosis, treatment and screening papers. Appendix 5 lists the search filters used.

Inclusion and exclusion criteria

Inclusion criteria

Review of tandem MS

Papers that fulfilled the following criteria were selected for the review and critically appraised:

- study design: all study types
- subjects: neonatal or newborn infants
- intervention: tandem MS
- outcome: data on (for the calculation of) the sensitivity, specificity or positive predictive value of screening.

Review of individual inborn errors of metabolism

For a defined list of inherited metabolic diseases detectable by tandem MS (*Table 2*), papers that fulfilled the following criteria were included in the review and critically appraised.

Studies of epidemiology

- study design: cohort studies and cross-sectional studies
- subjects: neonatal or newborn infants with an inherited metabolic disorder
- outcome: data on (for the calculation of) the incidence or prevalence of inborn errors of metabolism in the UK.

TABLE 2 *Included inherited metabolic diseases detectable by tandem MS*

Inherited metabolic disease
Phenylketonuria (PKU)
Tyrosinaemia type I
Homocystinuria
Maple syrup urine disease (MSUD)
Citrullinaemia (argininosuccinate synthase deficiency)
Ornithine carbamoyltransferase deficiency
Argininosuccinic aciduria (argininosuccinate lyase deficiency)
Arginase deficiency (argininaemia)
Hyperornithinaemia
Methylmalonic acidaemia
Propionic acidaemia
Isovaleric acidaemia
3-Hydroxy-3-methylglutaryl-CoA lyase deficiency
3-Methylcrotonyl-CoA carboxylase deficiency
Medium-chain acyl-CoA dehydrogenase (MCAD) deficiency
Carnitine palmitoyltransferase deficiency type I
Carnitine/acylcarnitine translocase deficiency
Carnitine palmitoyltransferase deficiency type II
Very long-chain acyl-CoA dehydrogenase deficiency
Trifunctional protein deficiency (long-chain 3-hydroxyacyl-CoA dehydrogenase deficiency)
Long-chain hydroxy-acyl-CoA dehydrogenase deficiency
Glutaryl-CoA dehydrogenase deficiency (GAI)
Multiple acyl-CoA dehydrogenase deficiency (GAII)

Studies of treatment outcome
- study design: randomised controlled trials (RCTs), cohort studies, case–control studies and case reports with more than two patients
- subjects: any subjects with an inherited metabolic disorder
- intervention: dietary, pharmacological, dialysis or surgical treatments
- outcome: any outcome of treatment.

Given that the purpose of this review was to establish that effective treatments exist and not to establish the effectiveness of all available treatment options, the review on the effectiveness of treatment outcomes was based on a best available evidence approach. In general, dietary and pharmacological treatments for inherited metabolic diseases require long-term management. For example, from diagnosis, the duration of treatment for PKU ranges from 6 years to lifetime;[1] therefore, studies of short duration were regarded as poor evidence for the review on the effectiveness of treatment outcome.

Exclusion criteria

Papers that fulfilled the following criteria were excluded from the review:

- subjects: non-human studies
- time: papers published before 1995 (this is an update of two systematic reviews,[1,2] that covered published literature before this date)
- language: any papers not available in the English language
- individual inborn errors of metabolism: non-UK data on the incidence or prevalence of inherited metabolic disorders and other methods of screening for individual inborn errors of metabolism, as the aim of this review is to evaluate the clinical effectiveness of neonatal screening using tandem MS
- other individual inborn errors of metabolism: the following inherited metabolic disorders, which may be detected by tandem MS in the neonatal period, were also excluded in this review: peroxisomal disorders (multiorgan diseases that are progressive and largely incurable),[1,2] biliary atresia (no accepted or specific screening method),[1] biotinidase deficiency (considered to be relatively rare in the UK and no direct experience of detecting the disorder using tandem MS),[1,2] 5-oxoprolinase deficiency and short-chain acyl-CoA dehydrogenase deficiency (uncertainty surrounding this condition and lack of accepted treatment).[1]

Literature identified

The yield of literature from the systematic searches by topic or disease is shown in *Table 3*. All citations and abstracts were assessed by one reviewer. Where there was insufficient information in the abstract, or where no abstract was available, a copy of the full paper was obtained and assessed for inclusion.

Data extraction strategy

All selected papers were read and appraised by a single reviewer, who extracted relevant information from the included studies directly into an evidence table. Any uncertainties were resolved by discussion with another reviewer and/or clinical advisers.

Quality assessment strategy

Checklists recommended by the NHS Centre for Reviews and Dissemination (CRD)[12] were used to assess the quality of systematic reviews, RCTs and non-randomised studies (e.g. cohort studies, case–control studies and case series), whereas diagnostic and screening studies were assessed using guidelines recommended by Greenhalgh and Donald[13] and Jaeschke and colleagues.[14,15] The overall strength of evidence included in the review was graded according to the levels of evidence as defined in the review by Pollitt and co-workers.[1] This has been summarised below.

Expected incidence (cases per 100,000) average within the UK

I Data obtained from whole population screening or comprehensive national surveys of clinically detected cases.
Ia As I but more limited in geographical coverage or methodology.
II Extrapolated from class I data for non-UK (but broadly similar) populations.
III Estimated from the number of cases clinically diagnosed in UK (compared with similar disorders of established incidence).

Effectiveness of treatment

The hierarchy of evidence follows standard guidelines published by the NHS CRD:[12]

I Well-designed RCTs.
II-1 Well-designed controlled trials with pseudo-randomisation or no randomisation.
II-2 Well-designed cohort studies.

TABLE 3 *Summary of the literature identified by systematic searching*

Disorder/technique	Literature search: no. of articles identified						
	No. of articles identified	Not relevant	Selected for further evaluation	UK incidence/ prevalence	Neonatal screening	Treatment	
						Identified for review	Included in review[a]
Tandem MS	145	91	54	NA	15	NA	NA
PKU	363	249	114	1	NA	9	7
Tyrosinaemia type 1	84	67	17	1	NA	2	1
Homocystinuria	194	161	33	0	NA	6	5
MSUD	59	37	22	0	NA	0	0
Urea cycle disorders: ornithine carbamoyltransferase (transcarbamylase) deficiency, citrullinaemia, argininosuccinic aciduria, arginase deficiency, hyperornithinaemia	147	125	22	0	NA	2	2
Methylmalonic, propionic and isovaleric acidaemia	99	62	37	0	NA	2	2
Other defects of branched-chain acyl-CoA metabolism: 3-hydroxy-3-methylglutaryl-CoA lyase deficiency, 3-methylcrotonyl CoA carboxylase deficiency	51	42	9	0	NA	0	0
MCAD deficiency	108	58	50	3	NA	0	0
Defects of long-chain fatty acid catabolism: carnitine palmitoyltransferase deficiency type I, carnitine/acylcarnitine translocase deficiency, carnitine palmitoyltransferase deficiency type II, very long-chain acyl-CoA dehydrogenase deficiency, trifunctional protein deficiency (long-chain 3-hydroxyacyl-CoA dehydrogenase deficiency), long-chain hydroxy acyl-CoA dehydrogenase deficiency	245	215	30	0	NA	0	0
GAI	97	76	21	0	NA	2	2
GAII	69	65	4	0	NA	0	0

[a] Best available evidence approach.
NA, not applicable.

(a) prospective with concurrent controls
(b) prospective with historical control
(c) retrospective with concurrent controls.

II-3 Well-designed case–control (retrospective) studies

III Large differences from comparisons between times and/or places with and without intervention (in some circumstances these may be equivalent to level I or II).

IV Opinions of respected authorities based on clinical experience, descriptive studies and reports of expert committees.

Sensitivity/specificity (%) of screening process for severe or classical disease, assuming 100% coverage

I Data obtained from screening programmes in the UK population or similar.

II Data from systematic studies other than from whole population screening.

III Estimated from known biochemistry of the condition.

Data synthesis

Formal meta-analytical techniques were not used for neonatal metabolic screening because of the concerns over heterogeneity and poor study quality expressed in the previous systematic reviews.[1] Instead, summary results are tabulated with detailed descriptive qualitative analyses. Sensitivities and specificities were reported to the nearest whole percentage in the review of neonatal screening using tandem MS (Chapter 5) and in the reviews of inborn errors of metabolism (Chapter 6), whereas in the Appendices (7–9), the sensitivities and specificities have been reported to three decimal places. For the economic models, false-positive and false-negative rates were calculated to the appropriate significant figure (Chapter 7, Economic modelling and Appendices 39–42).

Limitations of review methodology

Articles published after January 2002 were not included in the present review. The previous HTA publications sought up-to-date information (activity, resources, and research and development) by sending out a questionnaire to all newborn screening laboratories in the UK and site visits were made to laboratories in the USA and the UK to assess new methodologies for newborn screening (including tandem MS). However, this approach was beyond the scope of the current review. The authors were aware of various unpublished literature, conference proceedings and reports regarding the effectiveness of neonatal screening using tandem MS. Some of these have been cited throughout the report.

Because of the nature and size of the present review and the timetable for completion, it was not possible to contact all authors of papers to clarify ambiguities or fill in missing details from the (un)published literature. To identify all of the references related to the clinical and cost-effectiveness of neonatal screening for inborn errors of metabolism using tandem MS, searches were carried out without any language restrictions. At the data extraction stage, non-English language publications were excluded as the timescale of the review precluded time for translation. For each section, a bibliographic list of all excluded studies, including those excluded on language criteria, is provided in the appendices.

Chapter 4

Quantity and quality of evidence from existing systematic reviews

Evidence from the Pollitt report

All study types were included in the report by Pollitt and colleagues,[1] with no other inclusion criteria specified. Supplemental information was obtained from two laboratories in the UK and two from the USA where neonatal screening, based on tandem MS, was being developed for inborn errors of metabolism. The authors selected 28 studies for the review on tandem MS. Of these, eight were relevant for the evaluation of neonatal screening using tandem MS.

The authors found that "for the most part the quality of the literature uncovered had fallen far short of what one would normally require as the basis for a systematic review". However, limited information is provided regarding the methodological quality of identified studies, investigation of differences between studies and information on primary studies (e.g. study design and quality, patient characteristics, diagnostic methodology, outcome measures and sensitivity and/or specificity of the screening process). The number of included studies used to produce the summary of evidence is also unclear.

Based on the evidence of the literature and supplementary information from the laboratories, the authors concluded that "there appears to be a strong case for introducing tandem MS based screening. Screening should be limited to clearly defined diseases where specificity is known to be accurate and there are satisfactory confirmatory tests. Given the technical complexity of the method, and limited experience of applying tandem MS based screening to UK populations, a 3 year pilot study is proposed. …"[1]

Evidence from the Seymour report

The methods for study selection and appraisal were well described in the report by Seymour and colleagues.[2] The literature review was combined with supplemental information obtained from four laboratories (two from the USA and two from the UK), where tandem MS was being developed or applied to neonatal screening for metabolic disorders.

The authors did not report the number of studies that were included in the review of neonatal screening using tandem MS. Despite the critical appraisal of all selected papers using a standard checklist, limited information is provided regarding the methodological quality of identified studies, investigation of differences between studies and information on primary studies (e.g. study design and quality, patient characteristics, diagnostic methodology, outcome measures and sensitivity and/or specificity of the screening process). It is also unclear how many studies were used to summarise the evidence for each disorder or technology.

Based on findings from the comprehensive literature review and supplementary information from the laboratories, Seymour and colleagues[2] concluded that "there is however insufficient evidence at present for the widespread introduction of tandem MS technology into newborn screening programmes in the UK. Tandem MS for newborn screening for phenylketonuria, medium-chain acyl-CoA dehydrogenase deficiency and glutaric aciduria type I should be further evaluated by primary research conducted over 5 years. During this primary research, and until reports are presented and decisions made, there should be an embargo on the introduction of tandem MS technology into newborn screening laboratories in the UK."

Chapter 5

Review of neonatal screening using tandem MS

Background

Tandem MS with two mass spectrometers on either side of a collision cell offers a broad-spectrum approach to screening relatively large number of disorders with rapid throughput per sample (approximately 2 minutes)[16,17] and simple sample preparation. New technological advances such as automated sample introduction make the tandem MS technique a suitable method for population screening.[16]

The basis of the methodology generally involves a punched sample from the dried (Guthrie) blood-spot card, which is extracted with solvent containing appropriate standards labelled with stable isotopes. The extracted metabolites and standards are then converted into butyl esters and these are identified and quantified using tandem MS.[2] Further technical details of tandem MS are provided in the publication by Seymour and colleagues.[2]

Quantity and quality of research available

Quantity of research available

The yield of literature from the systematic searches generated 145 publications on neonatal screening using tandem MS between 1966 and November 2001. Full copies of 54 papers (inclusion of studies limited to 1995 onwards, as earlier studies had already been identified by the Pollitt[1] and Seymour[2] reviews) that appeared to be relevant by screening citations and abstracts were obtained. Of these 54 papers, 15 studies[4,5,7–9,16–25] (including two abstracts[4,5] that reported the same study as Wiley and co-workers,[16] but provided additional information or results on tandem MS-based newborn screening from different periods) were judged to be relevant to the update of the two existing systematic reviews of neonatal screening using tandem MS.[1,2]

Thirty-nine studies were excluded from the update review (details of references provided in Appendix 6) for the following reasons:

- patients: not neonatal or newborn (n = 1)
- screening method: not tandem MS (n = 2)
- publication type: foreign language publication (n = 3), letters or comments (n = 3) and review (n = 1)
- inborn error: peroxisomal disorder (n = 1) and biliary atresia (n = 2)
- study design: not screening or diagnosis (n = 1)
- outcome measures: no data for the calculation of sensitivity or specificity (n = 25).

Quality of included studies

Studies of amino acids and acylcarnitines (Appendix 7)

Seven studies evaluated neonatal screening of amino acids and acylcarnitines using tandem MS.[7–9,16–19] Of these seven studies, three were reported as abstracts and provided limited information.[8,9,18]

Newborn infants were screened in Australia (New South Wales Newborn Screening Programme),[16] Germany (Bavaria Newborn Screening Program),[18] Saudi Arabia,[17] Taiwan[19] and the USA (North Carolina Newborn Screening Program,[8] Wisconsin Newborn Screening Program[9] and the New England Newborn Screening Program).[7] The total number of infants screened in newborn screening programmes ranged from approximately 50,000[9] to 257,000,[7] whereas in non-newborn screening programmes (patients recruited from a secondary care setting) the samples ranged from 2100[19] to 27,624.[17] Only one of these studies reported that specimens were collected consecutively[16] and only one reported loss to follow-up.[17]

Of the seven included studies, four had a prospective cohort design with study durations from 2–3 years.[7,16–18] One study did not provide any details of study design;[8] however, the assessment phase lasted for 11 months. Two studies failed to report the study design and duration.[9,19]

Most newborn dried blood-spot specimens were obtained between 1[7,17] and 3[7,16] days after the infant's birth. One study did not report the sample type, but the sampling was done at 3 days of age.[18] In contrast, Lin and colleagues[19] analysed dried blood spots but failed to report the age of sampling. Two studies did not report the sample type or age of sampling.[8,9]

All neonatal samples were analysed using tandem MS. Three studies used ESI tandem MS,[16,17,19] one study used a Micromass LC triple quadropole tandem MS[7] and three studies did not report the type of tandem MS used.[8,9,18] Of these seven studies, four used automated samplers and computer-assisted software for automated processing and recognition analysis.[7,16–18]

Various thresholds (cut-off limits) were used to identify inherited metabolic disorders. Some used the 99.5th percentile as the upper cut-off limit[16,17] and 0.5th percentile as the lower cut-off limit for some key metabolites.[17] Others used concentrations of each compound or ratios as 4 standard deviations (SD) above the mean[9,19] or ≥ 5 SD for some metabolites.[9] Zytkovicz and colleagues[7] also used cut-off values for each marker, but these ranged from 6 to 13 SD above the mean. No cut-off values were reported by two studies.[8,18]

Diagnosis of inherited metabolic disorders was generally established using repeat analysis and/or repeat blood-spot specimens.[7,16,17] Other confirmation techniques included capillary electrophoresis, thin-layer chromatography,[16] gas chromatography/mass spectrometry (GC/MS) analysis of urine[17,19] and mutation analysis using DNA techniques.[7,16]

Amino acid and acylcarnitine disorders that were screened included: PKU,[7,16,17,19] tyrosinaemia,[7,16,19] homocystinuria (relies on detection of hypermethioninaemia),[7,8,19] MSUD,[7,16,17,19] citrullinaemia,[7,17,19] argininosuccinic lyase deficiency,[7,8] methylmalonic aciduria,[7,8,17,19] propionic acidaemia,[7,17,19] isovaleric acidaemia,[17,19] 3-hydroxy-3-methylglutaryl-CoA lyase deficiency,[7,19] 3-methylcrotonyl-CoA carboxylase deficiency,[7,18] short-chain acyl-CoA dehydrogenase deficiency,[7–9,16] MCAD deficiency,[7–9,16–19] carnitine uptake defect,[18] carnitine palmitoyltransferase deficiency type I,[18] carnitine acylcarnitine translocase deficiency,[7] carnitine palmitoyltransferase deficiency Type II,[7] very long-chain acyl-CoA dehydrogenase deficiency,[7,18,19] long-chain hydroxyacyl-CoA dehydrogenase deficiency,[7,18,19] glutaryl-CoA dehydrogenase deficiency,[7,8,16,17,19] glutaric aciduria type II,[7,19] 2-methyl aceto-acetyl-CoA thiolase (*β*-ketothiolase) deficiency,[7,16] hyperphenylalaninaemia (HPA)[7,8,16,19] and multiple carboxylase deficiency.[7,19]

The strength and quality of the evidence for the screening process of amino acids and acylcarnitines ranged from grade I[7–9,16,18] to grade II.[17,19]

Studies of MCAD deficiency only (Appendix 8)

Four studies assessed newborn screening for MCAD deficiency using tandem MS. These included studies provided data from newborn screening programmes in Australia (New South Wales Screening Programme)[20] and the USA (Neo Gen Screening, Pennsylvania and North Carolina Newborn Screening Program)[22] and from systematic screening studies from the UK (Northern region of the NHS)[21] and the USA (Pennsylvania, Ohio, New Jersey, Illinois, Florida and North Carolina).[23]

The total number of infants screened ranged from 100,600[21] to 930,078.[23] Only one of these studies reported that specimens were collected consecutively[20] and none of the studies reported loss to follow-up or ethnic characteristics. Of the four included studies, three used a prospective cohort design with study durations from approximately 2[20] to 7[23] years, whereas a study from the UK, of approximately 3 years, used a retrospective cohort design.[21]

Most newborn dried blood-spot specimens from Australia[20] and the USA[22,23] were obtained < 72 hours after birth. The UK[21] study, which analysed dried blood spots, failed to report the age at which the samples were taken. All neonatal samples were analysed using tandem MS. Two studies used ESI tandem MS[20,21] and two used isotope dilution liquid secondary ion tandem MS with automated samplers.[22,23]

Various thresholds were used to identify MCAD deficiency. The UK study[21] used concentrations of octanoylcarnitine > 0.3 μmol/l with a high octanoylcarnitine:hexanoylcarnitine ratio (> 4.0) as the cut-off value. These values were based on specimens taken from 18 neonates with MCAD deficiency. The Australian study[20] used a threshold of octanoylcarnitine concentration ≥ 1.0 μmol/l for prospective screening, based on a retrospective analysis of newborn screening samples from 13 patients (dried blood-spot samples were obtained between 4 and 6 days for 12 patients and at day 10 for one patient), later diagnosed clinically with MCAD deficiency, and prospective analysis of newborn screening samples from 24,000 newborns. In contrast, a study from the USA[22] used a cut-off value of octanoylcarnitine concentration ≥ 0.3 μmol/l. This cut-off limit was based on 113 randomly collected blood spots from healthy newborn babies and 16 blood spots from newborns confirmed with MCAD deficiency. The Andresen[23] study established a range of cut-off values to detect mild and severe forms of MCAD

deficiency. The cut-off value for mild profiles was characterised by octanoylcarnitine concentration between 0.5 and 2.0 μmol/l with an octanoylcarnitine:decanoylcarnitine ratio between 2 and 4, whereas a severe profile was defined by octanoylcarnitine concentrations > 2.0 μmol/l and an octanoylcarnitine:decanoylcarnitine ratio > 4.

Diagnosis of MCAD deficiency was generally established using repeat analysis and/or repeat dried blood-spot samples,[20,23] together with analysis of plasma or urine samples.[20,21] Mutation analysis was verified using DNA techniques[22,23] or the polymerase chain reaction (PCR) assay.[20] It should be noted that DNA analysis and PCR assays are not alternatives. PCR is a way of amplifying samples to allow DNA mutation analysis to be performed. Also, repeat testing may be done as a first pass before DNA analysis (Grosse S: personal communication; 2002).

The strength and quality of the evidence for the screening process of MCAD deficiency ranged from grade I[20,22] to grade II.[21,23]

Laboratory-based studies (Appendix 9)

Two retrospective, analytical studies were identified. A study from the USA[24] compared the results of previous analyses by fluorometry for PKU with those by tandem MS. Neonatal dried blood-spot samples from 208 newborns, < 24 hours of age, were retrieved from storage from the California Newborn Screening Program (original collection dates ranged from 1992 to 1994). All of the retrieved samples were analysed in a blinded fashion using isotope dilution secondary ion tandem MS.

Specimens that were originally analysed for HPA by fluorometry used a phenylalanine concentration ≥ 258 μmol/l as a positive cut-off value. Metabolic specialists made the final diagnoses of cases of classical PKU or variant PKU. In contrast, screening for PKU using tandem MS, a threshold limit of 180 μmol/l was used to identify increased phenylalanine levels. Furthermore, a phenylalanine:tyrosine ratio of 2.5 was used as an additional indicator of primary HPA. These cut-off values were based on findings from a previous study, which was conducted by these authors.

A study from Saudi Arabia[25] developed and assessed a computer-assisted metabolic profiling algorithm (CAMPA) for automated flagging of abnormal metabolic profiles of amino acids and acylcarnitines by automated ESI tandem MS. Dried blood-spot samples were taken from newborn babies from various hospitals throughout Saudi Arabia and from sick children referred to the metabolic unit at the King Faisal Specialist Hospital. The authors did not report the age at which samples were taken.

The reference range and cut-off values for the diagnostic parameters were established by measuring either metabolite concentrations or peak ratios of certain metabolite pairs from known cases of organic acidaemia and amino acid disorders and from control samples. The criteria for the control sample were based on 1100 newborn infants with a minimum birth weight of 2.01 kg and sample analysis in < 72 hours from the time of collection. Accordingly, a cut-off value using the 99.5th percentile was set as the upper cut-off limit and the 0.5th percentile as the lower cut-off limit for some key metabolites.

Two retrospective analyses, using 559 tandem MS data files and a larger data set of 1151 files, were carried out to verify the performance of CAMPA and to quantify the sensitivity and specificity of the algorithm method.

The strength and quality of the analytical evidence for the screening process of amino acids and acylcarnitines were grade II.[24,25]

Summary of results

A summary of the results, including evidence grades, is provided in *Table 4*.

Amino acids and acylcarnitines

The total number of specimens identified as abnormal ranged from 13 in 50,000[9] to 498 in 166,000.[18] In newborn screening programmes from Australia (New South Wales),[16] Germany (Bavaria)[18] and the USA (Wisconsin[9] and New England[7]), abnormal amino acid and acylcarnitine samples constituted less than 0.3% of the total samples screened. In comparison, abnormal samples ranged from 0.31%[17] to 1.38%[19] in non-newborn screening programmes. The slightly higher values from these studies may be attributable to the smaller number of newborns screened.

Very few studies reported false-negative results.[8] Some of the authors who reported false-negative results[16–18] were reluctant to use their early experience as a basis for accurate estimates of false-negative rates.

TABLE 4 *Effectiveness of neonatal screening using tandem MS*

Study	Screening programme, country	Total screened (*n*)	True positive (*n*)	False negative (*n*)	False positive (*n*)	True negative (*n*)	Sensitivity (%)[a]	Specificity (%)[a]	Positive predictive value (%)[a]	Quality of evidence (grade)
Studies of amino acids and acylcarnitines										
Hoffman *et al*.[9]	Wisconsin Newborn Screening Program, USA	50,000	5	NR	8	49,987	NA	100	38	I
Lin *et al*.[19]	Taiwan	2,100	2	NR	27	2,071	NA	99	7	II
Muenzer *et al*.[8]	Neo Gen Screening, Pennsylvania, USA (Pilot study)	194,384	31	NR	228	194,125	NA	100	12	I
	North Carolina Newborn Screening Program, USA	131,776	27	3	NR[b]	NR	90	NA	NA	
Rashed *et al*.[17]	Saudi Arabia	27,624	20	0	67	27,537	100	100	23	II
Roscher *et al*.[18]	Bavaria Newborn Screening Program, Germany	166,000	49	0	449	165,502	100	100	10	I
Wiley *et al*.[16]	New South Wales Newborn Screening Program, Australia	196,000	46	3	164	195,790	94	100	22	I
Zytkovicz *et al*.[7c] (range)	New England Newborn Screening Program, USA	*Amino acids* (164,000–257,000)	22 (0–18)	NR	238 (2–74)	(163,958–256,992)	NA	Cumulative 100	Cumulative 8 (22 of 260)	I
		Acylcarnitines (164,000–184,000)	20 (0–10)	NR	213 (1–42)	(163,964–183,948)	NA	Cumulative 100	Cumulative 9 (20 of 233)	
Studies of MCAD deficiency only										
Andresen *et al*.[23]	Pennsylvania, Ohio, New Jersey, Illinois, Florida, North Carolina, USA	930,078	62	NR	0	930,016	NA	100	100	II
Carpenter *et al*.[20]	New South Wales Newborn Screening Programme, Australia	275,653	12[d]	NR	11	275,630	NA	100	52	I
Chace *et al*.[22]	Neo Gen Screening, Pennsylvania, and North Carolina Newborn Screening Program, USA	283,803	16	0	0	283,787	100	100	100	I
Pourfarzam *et al*.[21]	Northern region of the NHS, UK	100,600	8	0	0	100,592	100	100	100	II
Zytkovicz *et al*.[7]	New England Newborn Screening Program, USA	184,000	10	NR	42	183,948	NA	100	19	I

continued

TABLE 4 *Effectiveness of neonatal screening using tandem MS (cont'd)*

Study	Screening method	Target condition(s)	Total screened (*n*)	True positive (*n*)	False negative (*n*)	False positive (*n*)	True negative (*n*)	Sensitivity (%)[a]	Specificity (%)[a]	Positive predictive value (%)[a]	Quality of evidence (grade)
Laboratory-based studies											
Chace *et al*.[24]	Fluorometry	PKU	203	19	0[e]	91	93	100	51	17	II
	Tandem MS		203	19	0[e]	3	181	100	98	86	
Rashed *et al*.[25]	CAMPA tandem MS	Amino acids	559	119	0	91	349	100	79	57	II
	CAMPA tandem MS	Acylcarnitines	1151	147	0	153	851	100	85	49	

[a] Rounded to the nearest whole percentage.
[b] The authors reported the false-positive rate as < 0.85%.
[c] Summary of results for this study reported in the amino acid and acylcarnitine section, however, this study also provided individual results for MCAD deficiency.
[d] Includes one probable mild case.
[e] Values not reported in original paper, assumed zero.
NR, not reported. NA, not applicable.

Data from only three newborn screening programmes (New South Wales,[16] Bavaria[18] and North Carolina, USA)[8] found that the sensitivity of neonatal screening for amino acids and acylcarnitines, using tandem MS, ranged from 90%[8] to 100%.[18] In contrast, a non-newborn screening programme study from Saudi Arabia, which assessed 27,624 newborn babies, showed that the sensitivity of neonatal screening using tandem MS was 100%.[17]

Newborns with false-positive results may be adversely affected by the risks associated with investigation of the screen-detected abnormality. False-positive values from neonatal screening programmes ranged from 8 in 50,000[9] to 449 in 166,000[18] specimens screened, with false-positive rates less than 0.3% [New South Wales 0.08%,[16] Bavaria 0.27%,[18] Wisconsin 0.02%[9] and New England aggregate false-positive rate for all disorders (amino acids and acylcarnitines) approximately 0.3%].[7] False-positive values were not reported for the North Carolina Newborn Screening Program;[8] however, these authors reported the false-positive rate as less than 0.85%. The range of false-positive values was smaller in non-newborn screening programmes and these ranged from 27 in 2100[19] to 67 in 27,624[17] specimens screened. In these studies, the false-positive rates were 1.29% and 0.24%, respectively. In addition, data from a pilot study of 194,384 neonatal blood samples, conducted by Neo Gen (Pennsylvania, USA) and reported by Muenzer and colleagues,[8] found 228 false positives, with a false-positive rate of 0.12%.

All seven studies provided information on specificity and positive predictive value. Results obtained from all the newborn screening programmes (New South Wales,[16] Bavaria,[18] Wisconsin[9] and New England[7]) showed that the proportion of individuals without the disease who were correctly identified by the screening process was 100%.[7,9,16,18] Similar findings were observed in non-newborn screening programmes. Data from Taiwan and Saudi Arabia suggest specificities ranging from 99%[19] to 100%,[17] and results from a pilot study, conducted by Neo Gen, showed that the specificity of screening amino acids and acylcarnitines using tandem MS was 100%.[8]

The positive predictive value, a further measure of assay performance, for the New South Wales,[16] Bavaria,[18] and Wisconsin[9] newborn screening programmes ranged from 10%[18] to 38%.[9] Results from the New England Newborn Screening Program, reported by Zytkovicz and colleagues,[7] found that the positive predictive value for all amino acid disorders was 8%; however, this predictive value was increased to 14% using a combination of individual markers and ratios. The positive predictive value for all of the acylcarnitine disorders identified was 9%. Similar results were observed from non-newborn screening programmes from Saudi Arabia[17] (23%) and Taiwan[19] (7%).

The range and total number of amino acid and acylcarnitine disorders that were detectable in newborns by using tandem MS are shown in *Table 5*.

MCAD deficiency only

Four studies evaluated newborn screening for MCAD deficiency using tandem MS.[20–23] One additional study,[7] which was included in the amino acid and acylcarnitine section, also provided individual data for MCAD deficiency. These findings have been incorporated into the table of results for MCAD deficiency (*Table 4*).

The total number of specimens identified as abnormal ranged from 8 in 100,600[21] to 62 in 930,078.[23] These abnormal samples, suggestive of MCAD deficiency, accounted for less than 0.03% of the total samples screened.

None of the prospective studies reported false-negative results, except for one study, which used combined quantitative data from prospective newborn screening of 283,803 infants by Neo Gen Screening, Pennsylvania, and the North Carolina Newborn Screening Program, USA. The authors of this study reported that no known false-negative results had been found. Moreover, none of the prospective studies included a rigorous method to identify those who might have been missed by the screening process. The UK retrospective study[21] did not identify any false-negative results after examination of the regional registers for metabolic diseases and deaths.

Only one prospective screening study, reported by Chace and colleagues,[22] found that the sensitivity of neonatal screening at Neo Gen Screening and North Carolina Newborn Screening Program, for MCAD deficiency, using tandem MS was 100%. In the UK retrospective study,[21] the sensitivity of the screening process was also found to be 100%; however, the authors reported that the sensitivity of the test was difficult to ascertain, because many occurrences of MCAD deficiency were not diagnosed on clinical grounds.

False-positive values from the prospective screening studies ranged from 0 in 283,803[22] to

TABLE 5 *Amino acid and acylcarnitine disorders detected by tandem MS*

Study:	**Hoffman et al.**[9a]	**Lin et al.**[19]	**Muenzer et al.**[8b]	**Muenzer et al.**[8c]	**Rashed et al.**[17]	**Roscher et al.**[18d]	**Wiley et al.**[16e]	**Zytkovicz et al.**[7f]
Total number of cases detected:	**(*n* = 5)**	**(*n* = 2)**	**(*n* = 31)**	**(*n* = 27)**	**(*n* = 20)**	**(*n* = 49)[g]**	**(*n* = 46)**	**(*n* = 42)**
Disorders of amino acid metabolism								
PKU					3		28	7
HPA		1	6	9			3	11
Biopterin defects							2	
Hypermethioninaemia			1					1
Tyrosinaemia type II							1	
Argininaemia								1
Argininosuccinase lyase deficiency			1		2			1
Citrullinaemia			3		1			
GAI					2		1	
MSUD					2		1	1
Disorders of organic acid metabolism			5[h]	7[h]				
Methylmalonic aciduria					4			
Propionic acidaemia					1			2
Isovaleric acidaemia		1			2			
3-Methylcrotonyl-CoA carboxylase deficiency								1
2-Methylacetoacetyl-CoA thiolase (β-ketothiolase deficiency)							1	
Disorders of fatty acid metabolism								
Short-chain acyl-CoA dehydrogenase deficiency	1			1			1	5[i]
Medium-chain acyl-CoA dehydrogenase deficiency	4		14	10	2	13	6	10
Very long-chain acyl-CoA dehydrogenase deficiency								1[i]
Multiple acyl-CoA dehydrogenase deficiency								
Carnitine uptake defect								
Carnitine palmitoyltransferase type I								
Carnitine palmitoyltransferase type II								1
LCAD deficiency			1					
LCHAD deficiency								

continued

TABLE 5 *Amino acid and acylcarnitine disorders detected by tandem MS (cont'd)*

Study:	**Hoffman et al.**[9a]	**Lin et al.**[19]	**Muenzer et al.**[8b]	**Muenzer et al.**[8c]	**Rashed et al.**[17]	**Roscher et al.**[18d]	**Wiley et al.**[16e]	**Zytkovicz et al.**[7f]
Total number of cases detected:	**(*n* = 5)**	**(*n* = 2)**	**(*n* = 31)**	**(*n* = 27)**	**(*n* = 20)**	**(*n* = 49)[g]**	**(*n* = 46)**	**(*n* = 42)**
Other disorders								
Non-ketotic hyperglycinaemia					1			
Vitamin B_{12}-deficient babies of vegan mothers							2	

[a] Wisconsin Newborn Screening Program, USA.
[b] Neo Gen Screening, Pennsylvania, USA (pilot study).
[c] North Carolina Newborn Screening Program, USA.
[d] Bavaria Newborn Screening Program, Germany.
[e] New South Wales Newborn Screening Programme, Australia.
[f] New England Newborn Screening Program, USA.
[g] Not all cases detected for each disorder are reported; however, the first cases of long-chain hydroxyacyl-CoA dehydrogenase deficiency and carnitine uptake defect and carnitine palmitoyltransferase type I were detected in prospective newborn screening. 3-Methylcrotonyl-CoA carboxylase deficiency (three isolated cases and two asymptomatic mothers) appeared to be the most frequent organic acid disorder. Other defects included severe variants of very long-chain acyl-CoA dehydrogenase deficiency and multiple acyl-CoA dehydrogenase deficiency (two newborns were identified in newborn screening, but died, on day 3, before diagnosis).
[h] Type of organic acid disorder not specified.
[i] Presumptive value.
LCAD: long-chain acyl-CoA dehydrogenase; LCHAD: long-chain hydroxyacyl-CoA dehydrogenase.

42 in 184,000[7] specimens screened, with false-positive rates less than 0.02% (New South Wales Screening Programme 0.004%,[20] Neo Gen Screening, Pennsylvania and North Carolina Newborn Screening Program 0%,[22] New England Newborn Screening Program 0.023%,[7] and from a systematic study from the USA – Pennsylvania, Ohio, New Jersey, Illinois, Florida and North Carolina – 0%).[23] It is noteworthy that in one study from Australia,[20] 11 infants were identified as false positives; however, four of these patients died from a variety of causes in the neonatal period before a second sample could be collected. MCAD deficiency was eliminated by enzyme analysis in cultured skin fibroblasts in one patient, whereas MCAD deficiency in the other three infants was eliminated on the basis of information obtained from clinicians and post-mortem findings. No false-positive results were found in the UK retrospective study.[21]

All five studies provided information on specificity and positive predictive value. Results obtained from the prospective newborn screening programmes (New South Wales Screening Programme,[20] Neo Gen Screening, Pennsylvania and North Carolina Newborn Screening Program,[22] and the New England Newborn Screening Program[7]) showed that the specificity of neonatal screening for MCAD deficiency, using tandem MS, was 100%.[7,20,22] Similar findings were observed in systematic screening studies from the UK (100% specificity)[21] and the USA (100% specificity).[23] The positive predictive value in the UK retrospective study was 100%,[21] whereas the positive predictive value for prospective newborn screening ranged from 19%[7] to 100%.[22,23]

Laboratory-based studies

Chace and colleagues[24] observed a strong correlation between the results obtained by fluorometry and tandem MS (Pearson correlation coefficient 0.817). In samples collected from newborns < 24 hours of age, phenylalanine measurements by tandem MS, with a cut-off of 180 μmol/l, or by fluorometry, with a cut-off of ≥ 258 μmol/l, detected all infants with confirmed variants of hyperphenylalaninaemia and classical PKU.

Tandem MS analysis greatly reduced the number of false-positive results, from 91 to three. Simultaneous quantification of phenylalanine and tyrosine by tandem MS further reduced the number of false positives to one, using a cut-off of phenylalanine:tyrosine molar ratio ≥ 2.5. The proportion of individuals without PKU who were correctly identified by the tandem MS analytical method was 98%. In contrast, the specificity was 51% using the fluorometric method of analysis. The positive predictive values were 86% and 17%, respectively.

Rashed and colleagues[25] carried out two retrospective experiments to determine the performance of a CAMPA for automated flagging of abnormal metabolic profiles of amino acids and acylcarnitines in newborns using tandem MS. These authors found that the sensitivity of the algorithm was 100%, collectively, for flagging cases with known metabolic disorders, and the average cumulative specificity for the two tests was 83%, with a positive predictive value ranging from 49 to 57%.

Discussion

The evidence for neonatal screening of amino acids and acylcarnitines using tandem MS is primarily from observational data of large-scale prospective newborn screening programmes, from several centres outside the UK, namely, Australia,[16] Germany[18] and the USA.[7–9] RCTs of screening for rare disorders are difficult because of the enormous numbers that would be needed for adequate power.[1,4,20]

The evidence is limited regarding the sensitivity (false negatives) and specificity (false positive) for each of the diagnosable disorders detected by tandem MS. Only one study provided data for calculation of sensitivities and specificities for individual amino acid and acylcarnitine disorders. In general, neonatal screening for the amino acid and acylcarnitine group of disorders using tandem MS demonstrated high sensitivities[8,16–18] and high specificities.[7–9,16–19]

More up-to-date information regarding the sensitivity and specificity for each of the diagnosable disorders detected by tandem MS has not been published yet. Major initiatives using tandem MS in neonatal screening programmes have started in several centres around the world. Those with the most thorough evaluation programmes are in New South Wales (Australia), New England (USA) and Bavaria (Germany). Other programmes are Neo Gen based in Pittsburgh and in North Carolina (USA), but additional programmes are being developed rapidly and many more have just started or will start very soon (Leonard J: personal communication; 2002). Data from some of these

ongoing programmes will become available when the results are known and the sensitivities and specificities of neonatal screening for individual inborn errors of metabolism will begin to be much more clearly defined.

Unpublished data from Heidelberg, Germany[26] [tandem MS-based neonatal screening of 250,000 neonates performed during an observation period of 41 months: amino acidaemias: sensitivity 92.75%, specificity 99.90% and positive predictive value 20.00%; fatty acid oxidation disorders: sensitivity 100%, specificity 99.90% and positive predictive value 9.29%; organic aciduria: sensitivity 100%, specificity 99.88% and positive predictive value 5.17%; resulting in a sensitivity of 95.54%, specificity 99.67% and positive predictive value of 11.52% for the total extended (41 months) screening], and Bavaria, Germany (3-year prospective tandem MS-based newborn screening study of 307,676 neonates: false-positive rate <0.45%; false-negative, only one case with tyrosinaemia type I identified so far) suggest that the results are very positive and in 2002, tandem MS-based neonatal screening became 'standard medical care' and is covered by the German state health scheme (Pollitt R: personal communication; 2002). Unpublished data from the New South Wales Newborn Screening Programme, Australia, suggest that tandem MS-based neonatal screening of 343,244 babies between April 1998 and December 2001, for selected amino acids (citrulline, glycine, leucine, methionine, phenylalanine and tyrosine) and acylcarnitines (C2, C3, C4, C5, C5OH, C5D, C6, C8, C10, C10:1, C14, C14:1, C16 and C16OH) provided a sensitivity of 93%, specificity of 99.8% and positive predictive value of 12%.[27]

Difficulties in the detection and differentiation of certain disorders, such as transient tyrosinaemia,[7,17] argininosuccinic lyase deficiency and[7,19] glutaryl-CoA dehydrogenase deficiency,[7] or metabolites, such as propionylcarnitine to acylcarnitine ratio,[8,17] 3-hydroxy-isovalerylcarnitine,[8] C5-carnitine[17] and tyrosine,[8] aided false-positive results. In contrast, the true number of false negatives will only be learned after newborn screening is implemented, and children who are not detected as newborns are diagnosed later in life. Wilcken and colleagues[5] who provided additional data from the New South Wales Screening Programme, concluded that newborn screening using tandem MS can reliably detect some inborn errors of metabolism such as phenylketonuria, many organic acidaemias and classic MSUD with confidence, whereas almost all patients with MCAD deficiency (who would later have developed with symptoms) and glutaryl-CoA dehydrogenase deficiency are detected with certainty. Disorders that could not be reliably detected included mild MSUD, homocystinuria, tyrosinaemia type I and probably some but not all fatty acid oxidation disorders. However, there is no justification in not seeking these mild or late-onset variants of metabolic disorders simply because they cannot be detected reliably on every occasion.[5]

Various thresholds were used to identify inborn errors of metabolism. Some studies used cut-off limits based on percentiles,[16,17] whereas others used various concentrations of each compound or ratio as standard deviations above mean.[7,9,19] Furthermore, infants and young children who are ill for any reason may have abnormal patterns of amino acids and acylcarnitines. Zytkovicz and colleagues[7] found that babies in neonatal intensive care units and very low birth weight infants (< 1500 g) had different marker profiles, indicative of inherited metabolic disorders, to those of normal birth weight infants. The diversity in the preference of metabolite, together with the cut-off limits used to define a positive outcome, facilitate the yield of false-positive and false-negative results and therefore restrict direct comparison of the performance of tandem MS between studies.

Worldwide, there is an increasing trend towards discharging mother and baby within the first day or two of life;[16] therefore in countries such as Australia,[16] Saudi Arabia[17] and the USA,[7] the age of sampling for selected amino acids and acylcarnitines is usually performed within 72 hours after birth, whereas in the UK, blood samples are usually taken at about 6–14 days of age.[1,2] The age at which screening is undertaken will affect the sensitivity and specificity of the screening process as concentrations of metabolites change over time. For example, in the UK, specimen collection is delayed until the sixth day of life or later to maximise sensitivity for the PKU screen; however, this time-lapse may be detrimental for disorders that present acutely in infancy, such as MSUD.[1]

Despite the high sensitivities[21,22] and high specificities[7,20–23] of neonatal screening for MCAD deficiency using tandem MS, the quality of the evidence restricts direct comparison about diagnostic performance and the outcome of tandem MS between studies. Prospective studies of MCAD deficiency[7,20,22,23] have used various analytes and thresholds for detecting disease

status. More importantly, in two studies from the USA,[22,23] infants were considered to have MCAD deficiency solely on the basis of diagnostic acylcarnitine profiles, whereas Carpenter and colleagues[20] and Zytkovicz and co-workers[7] applied explicit criteria for the diagnosis of MCAD deficiency. In the UK retrospective study,[21] which used explicit criteria for the diagnosis of MCAD deficiency, the authors reported that in most cases of MCAD deficiency, diagnosis was not based on clinical grounds.

Most dried blood-spot samples obtained in prospective newborn screening studies of MCAD deficiency were collected less than 72 hours after birth,[7,20,22,23] which is considerably earlier than in the UK, where neonatal screening samples are normally collected between 6 and 14 days of life.[1,2] To evaluate the influence of patient age on the concentrations of acylcarnitines in MCAD deficiency, Chace and colleagues[22] found that the diagnostic acylcarnitines (hexanoyl-, octanoyl-, decenoyl- and decanoylcarnitine) were higher in the newborn period (< 72 hours after birth) than observed in older patients (ages 8 days to 7 years) with MCAD deficiency. These authors concluded that the higher metabolite concentration in the newborn period facilitated the detection of MCAD deficiency. Clayton and co-workers,[28] who reported their experience in diagnosing MCAD deficiency in the UK population using the technique of ESI tandem MS analysis of butylated carnitine species from dried blood spots, concluded that if neonatal screening were undertaken at 7–10 days of age, the number of false-positive and false-negative results should be negligible.

New technological approaches for automated processing, coupled with computer-assisted software to provide individually established cut-offs and ratios of various analytes, have demonstrated tandem MS methodology to be a robust, highly sensitive and specific method[25] for the detection of inborn errors of metabolism. This approach has been adopted by the New South Wales Screening Programme, Australia,[16] the Bavaria Newborn Screening Program, Germany[18] and the New England Newborn Screening Program, USA.[7] This approach also allows the analysis of hundreds of samples on a daily basis and minimises labour costs.[16]

In summary, tandem MS has the potential for the simultaneous detection of a wide range of metabolic disorders using a single analytical technique. This method has been shown to be rapid, highly sensitive and highly specific for newborn screening of dried blood spots for the amino acid and acylcarnitine group of disorders. However, evidence is limited regarding the false negatives and false positives for individual diagnosable disorders. The variation in the age of sampling and the heterogeneity in the choice of metabolite, as well as in thresholds used to define a positive result, limits direct comparison of the discriminative power of tandem MS between studies.

Chapter 6

Reviews of inborn errors of metabolism

Phenylketonuria

Background

PKU is an inborn error of metabolism resulting from a deficiency of phenylalanine hydroxylase, which converts the amino acid phenylalanine to tyrosine. Without this enzyme, abnormally high concentrations of phenylalanine accumulate in blood and tissues. As a consequence, there is excessive production of phenylpyruvic, phenylacetic and phenyllactic acids (phenylketones) in tissues, resulting in high levels in the urine.[1]

The disorder is characterised by HPA and if untreated will usually lead to severe mental retardation, seizures and other neurological abnormalities. An individual with PKU usually appears normal in the first few months of life but signs of delayed development will be apparent by 6–12 months. Early dietary intervention, by dietary restriction of phenylalanine, prevents progressive, irreversible cognitive damage but does not reverse pre-existing damage.[2]

A number of variant forms of PKU also exist, including patients with mild to moderate HPA who may or may not require dietary treatment and biopterin cofactor variant forms that require treatment with tetrahydrobiopterin.[29] Classical PKU is the severe form of the disease, resulting from a complete deficiency of phenylalanine hydroxylase.[1]

Quantity and quality of research available

Quantity of research available

The systematic searches identified 363 recent publications on the incidence, prevalence and effectiveness of treatment outcomes for PKU between 1995 and January 2002 (limited to 1995 onwards, as earlier studies would have already been identified by the existing systematic reviews[1,2]). Full copies of 114 papers were retrieved for further evaluation.

Of these 114 papers, eight studies[30–37] were judged to be relevant to the update review. Only one publication provided data on the incidence of PKU in the UK. Based on the best available evidence approach (studies of short duration were regarded as poor-quality evidence for the review on the effectiveness of treatment outcome), two systematic reviews of RCTs and five additional studies provided information on the effectiveness of treatment outcomes for PKU.

One-hundred and six studies were excluded from the update review (details of references are provided in Appendix 10), for the following reasons:

- patients: not PKU (n = 1), prenatal (n = 1)
- publication type: letters, comments or statements (n = 8), review (n = 17)
- study design: screening (n = 18)
- outcome measures: mutation, genotype, etc., analysis (n = 18), no incidence or prevalence data (n = 2), no treatment outcomes, details on dietary treatment or fewer than two patients (n = 13), incidence or prevalence data from non-UK countries (n = 17)
- other reasons: same results reported elsewhere (n = 3), included in existing or selected systematic reviews (n = 6), poor-quality evidence: treatment duration < 48 hours (n = 2).

Quality of included studies

Incidence of PKU (Appendix 11)

Only one study reported data on the incidence of PKU in the UK.[30] This retrospective cohort study evaluated all neonatal infants (n = 707,720) born in the West Midlands NHS region (covering the areas of the West Midlands, Hereford, Worcester, Shropshire, Staffordshire and Warwickshire) during the period 1981–91 for inborn errors of metabolism in different ethnic groups (north-west European, Pakistani, Indian, Afro-Caribbean, other ethnic groups and mixed race). Neonatal samples for PKU screening were obtained 6–10 days after birth.

Retrospective data for inherited metabolic disorders were derived from the West Midlands Neonatal Screening Programme regional register for patients with inborn errors of metabolism, and population frequencies from the national census. All confirmations were obtained from laboratory records.

The Neonatal Screening Service programme for Birmingham used heparinised plasma and screened for PKU by amino acid chromatography, as opposed to dried blood spots and the Guthrie microbiological assay for phenylalanine, used by the rest of the region. The authors failed to report any details on gender characteristics and thresholds used for disease identification.

The strength and quality of the evidence for the expected incidence of phenylketonuria within the UK were grade Ia.[30]

Effectiveness of treatment for PKU
Two Cochrane systematic reviews of RCTs and five additional studies were identified that evaluated the effectiveness of treatments for PKU.

Dietary interventions for PKU (Appendix 12) One systematic review evaluated dietary interventions for PKU. In this review Poustie and Rutherford[31] performed a meta-analysis from published and unpublished clinical trials to assess the effect of a phenylalanine-restricted diet started early in life for patients with PKU and possible adverse effects of relaxation or termination of diet on intelligence, neuropsychological performance, effect on growth, nutritional status, eating behaviour, quality of life and death.

Original articles were assessed for inclusion in the meta-analysis on the following predetermined criteria: randomised or pseudo-randomised trials comparing a phenylalanine-restricted diet to either relaxation or termination of dietary restrictions in patients of any age with PKU. Comprehensive searches were conducted on various electronic searches (MEDLINE, EMBASE) between 1966 and January 2001, and handsearching journals and abstracts of conferences, reference lists and the Cystic Fibrosis and Genetic Disorders trials register. Manufacturers of very low or phenylalanine-free protein supplements were also contacted for published and unpublished RCTs on file.

The results of four trials were pooled for further analysis (n = 251) and weighted mean differences calculated. The majority of included studies involved only small numbers of subjects, ranging from nine to 216. In two studies, the mean age of the patients ranged from 12.6 to 14.4 years, but was unreported in the other included studies; however, treatment in these studies was initiated within 121 days after diagnosis. Moreover, several included studies failed to provide details on allocation concealment and method of randomisation sequence. Only two of the four included studies used intention-to-treat analysis. The length of follow-up was different in the various studies and ranged from 10 weeks to 12 years. The authors did not report the number of patients lost to follow-up for any of the studies.

More than 30 different assessments of neuropsychological performance were evaluated in the four included studies. Of these different assessments, only three were used in more than one study; however, final data from these assessments were not available. In addition, the following outcome measures were not measured in any of the studies: blood tyrosine concentration, eating behaviour, quality of life and mortality.

Two further non-randomised studies provided additional information on dietary interventions for PKU. The first study[35] assessed the effect of a phenylalanine-restricted diet in ten young PKU children (median age 7.53 years, range 5.58–9.75 years) versus diet termination in ten older PKU patients (median age 20.59 years, range 13.58–28.42 years) using a cross-sectional mixed design with matched pairs (controls free from major developmental disorders), independent samples and non-repeated measures. All PKU patients were diagnosed with classical PKU and treated within 3 months of birth.

The second study[36] evaluated the effect of different degrees of dietary intervention on vitamin B_{12} and folate status using a prospective cohort design with historical controls. Diseased patients included adolescents and adults with PKU on a strict low-phenylalanine diet with amino acid, mineral and vitamin supplements (n = 22, median age 24 years), relaxed diet (total protein intake of approximately 1 g/kg per day with roughly 50% from natural protein and 50% from amino acid, mineral and vitamin supplements, n = 30, median age 21 years) or unrestricted diet (no formal protein restriction and not taking amino acid supplements, n = 31, median age 22 years).

Neither non-randomised study reported any information on the period of study, inclusion and exclusion criteria, ethnic characteristics or loss to follow-up.

Tyrosine supplementation for PKU (Appendix 13) Only one study was identified on tyrosine supplementation for PKU. In this study, Poustie and Rutherford[32] performed a meta-analysis from published and unpublished clinical trials to assess the effects of tyrosine supplementation alongside or instead of a phenylalanine-restricted diet for

patients with PKU who started the diet on diagnosis and either continued on the diet or relaxed the diet later in life. The effects of tyrosine supplementation alongside or instead of a phenylalanine-restricted diet on intelligence, neuropsychological performance, growth and nutritional status, mortality rate and quality of life were also evaluated.

Original articles were assessed for inclusion in the meta-analysis on the following predetermined criteria: randomised or pseudo-randomised trials comparing the use of tyrosine supplementation versus placebo in patients with PKU in addition to, or instead of, a phenylalanine-restricted diet. Patients treated for maternal PKU were excluded. Comprehensive searches were conducted on various electronic searches (MEDLINE, EMBASE) between 1966 and June 2001, and handsearching journals and abstracts of conferences, reference lists and the Cystic Fibrosis and Genetic Disorders trials register. Manufacturers of dietary products used in the treatment of PKU were also contacted for published and unpublished RCTs on file.

The results of three trials were pooled for further analysis (n = 56) and weighted mean differences were calculated. Three included studies were double blind and placebo controlled with adequate allocation concealment. In all three studies, only small numbers of subjects were involved (range 9–24, with a mean age range of 10.25–20.08 years) and the duration of both the treatment and control arms were brief (range 12–24 weeks). Two studies failed to provide details of the method of randomisation sequence and only one of the three included studies used intention-to-treat analysis.

The intelligence quotient (IQ) was not assessed because the results of all the patients were combined and compared with a non-PKU group. In addition, the following outcome measures were not measured in any of the studies: weight gain, other measures of nutritional status, quality of life and death.

Fatty acid supplementation for PKU (Appendix 14) Two studies were identified that investigated the effects of fatty acid supplementation in children with either PKU or HPA.

In one study[33] the authors used a single blind randomised controlled design and recruited a group of treated PKU children (monitored for clinical symptoms and nutritional follow-up), of both genders, between 5 and 10 years of age. Over a period of 6 months, subjects underwent dietary supplementation with fish oil containing long-chain polyunsaturated fatty acids (LCPUFA) of the n-3 series (n = 10) or blackcurrant seed oil containing compounds of both n-6 and n-3 series immediately beyond the limiting step for LCPUFA synthesis (n = 11). Subjects received 2.5–4 g oil per day on the basis of individual body weight (one 500 mg capsule/4 kg body weight per day for 6 months). Treatment was allocated using a time-balanced randomisation table and no subjects were lost to follow-up. No details were provided on ethnic characteristics or power calculations for detecting differences.

Agostoni and colleagues[34] used a double-blind randomised placebo-controlled trial. Children included in this study were detected by newborn screening and diagnosed as having type I HPA according to predefined protocols. Twenty-four children with HPA, of both genders, treated from the first month of life, were randomly allocated to receive gelatine capsules containing 500 mg oil per capsule supplying either 26% fatty acid as LCPUFA (including 4.6% γ-linolenic acid, 7.4% arachidonic acid, 5.5% eicosapentaenoic acid and 8% docosahexaenoic acid) or placebo (olive oil) supplements over a 12-month period. The daily dosage of the supplement provided 0.3–0.5% of the daily energy requirements as LCPUFA or an equivalent dose as placebo (around 1 capsule/4 kg body weight). No details were provided on the method of randomisation sequence or ethnic characteristics.

Based on power calculations, nine patients were required per group for detecting differences at 90% power. Anticipating a 25% loss to follow-up over the study period, 12 children were recruited for each group. The mean age of the patients in the treatment group was 10 ± 7 years and 10 ± 5 years in the placebo group. Two patients in each group were lost to follow-up owing to parental withdrawal.

Treatment during pregnancy (Appendix 15) One study was identified that determined whether dietary treatment during pregnancy of women with PKU affected developmental outcomes of offspring.[37] This ongoing longitudinal, prospective cohort study (Maternal PKU Collaborative Study), which began in 1984, followed a total of 253 children of women with PKU (n = 149 children, further subdivided into four treatment groups related to the timing of maternal metabolic control: control before pregnancy, n = 17; more than 0 up to 10 weeks of gestation, n = 26; more than 10 up to 20

weeks of gestation, 47; more than 20 weeks of gestation or never in control, $n = 59$), with untreated mild HPA ($n = 33$) or without known metabolic problems (comparison group, $n = 71$), up to 4 years of age.

Women with PKU were offered a low-phenylalanine diet before or during pregnancy, with the aim of maintaining metabolic control (plasma phenylalanine ≤ 605 μmol/l), whereas women with mild HPA, who had plasma phenylalanine < 605 μmol/l on a normal diet, were not treated. No details were reported on baseline comparability, ethnic characteristics or loss to follow-up.

Overall, the strength and quality of the evidence for the effectiveness of treatments for PKU included four grade I studies,[31–34] one grade II-2a,[37] one grade II-2b[36] and one grade III[35] study.

Summary of results

Incidence of PKU

Study details and results are summarised in *Table 6*.

The overall incidence of PKU (classical and atypical combined) in the West Midlands NHS region of the UK was 7.3 cases per 100,000 births between 1981 and 1991.[30] The incidence of PKU was similar in Pakistani [1:14,452, 95% confidence interval (CI) 1:4001 to 1:119,337] and Caucasian children (1:12,611, 95% CI 1:9512 to 1:17,104); however, the gene frequency was significantly lower in Pakistanis (1:713, 95% CI 1:210 to 1:5750) than among north-west Europeans (1:112, 95% CI 1:98 to 1:131, $p < 0.01$). Two cases of PKU were identified in mixed-race children. One case was identified in a mixed-race Jordanian–European and one of Afro-Caribbean–Arabic descent. No cases of PKU were observed among Indians.

The authors reported that the incidence of inborn errors of metabolism in Afro-Caribbean populations in the West Midlands was not compared with that in north-west Europeans because the Afro-Caribbean population was too small.

Effectiveness of treatment for PKU

Dietary interventions for PKU

Data from a meta-analysis of four RCTs[31] found that blood phenylalanine concentrations were significantly lower in subjects with PKU following a phenylalanine-restricted diet than in those on a less restricted or relaxed diet [weighted mean difference at 3 months –672.203 (95% CI –813.799 to –530.608, $p = 0.00$); weighted mean difference after 12 months –751.540 (95% CI –883.412 to –619.677, $p = 0.00$)], as would be expected. IQ was significantly higher in subjects who continued on the phenylalanine-restricted diet than in those who terminated the diet [weighted mean difference after 12 months –5.000 (95% CI –9.595 to –0.405, $p = 0.03$)]; however, these findings were based on only one study.

Results from a cross-sectional study[35] showed that neuropsychological tests failed to provide compelling evidence that prolonged exposure to unrestricted phenylalanine during adolescence and early adulthood is harmful to cognitive and motor functioning in well-treated PKU patients, and delaying dietary termination until 10 years of age is sufficient to prevent a substantial reduction in cognitive and motor ability. However, there were indications that subtle intellectual deficiencies may arise during and after treatment, possibly in the realm of frontal-executive functions. Inferences to wider PKU populations should be made with caution owing to small sample sizes and limited function tests.

In a prospective cohort study, Robinson and co-workers[36] showed that mean vitamin B_{12} levels were significantly lower in the PKU groups on relaxed or unrestricted diets than in the normal population (strict low-phenylalanine diet: 468.7 ± 199.7 ng/l, $p = 0.077$ vs normal population; relaxed diet: 332.8 ± 128 ng/l, $p = 0.0034$ vs normal population; unrestricted diet: 275.3 ± 95 ng/l, $p = 0.0001$ vs normal population; normal population: 411.9 ± 148.75 ng/l). The results illustrate that adolescents and adult patients with PKU who have discontinued or relaxed their diet might be at risk from vitamin B_{12} deficiency.

Tyrosine supplementation for PKU

Data from a meta-analysis of three RCTs[32] showed that the blood tyrosine concentrations in subjects with PKU were significantly higher in subjects receiving tyrosine supplements than those in the placebo group [weighted mean difference at 3 months 22.996 (95% CI 12.895 to 33.097, $p = 0.00$)]. No other significant differences ($p < 0.05$) were found in any of the other outcomes measured (blood phenylalanine and neuropsychological performance).

Fatty acid supplementation for PKU

Two RCTs provided data on the effectiveness of fatty acid supplementation in children with either PKU or HPA.

Agostoni and colleagues[33] reported that in children with PKU, aged 5–10 years, fish oil supplementation

TABLE 6 *Birth incidence of PKU in the UK*

Study	Location	Study design	Period of study	Patient selection	Age at sampling	Confirmation of disease	Population screened	Cases diagnosed	Cumulative incidence	Cases[a] per 100,000	Quality of evidence (grade)
Hutchesson *et al.*[30]	Neonatal Screening Service Programme, West Midlands, UK	Retrospective cohort study	April 1981 to April 1991	All newborns within West Midlands NHS region	6–10 days	Yes, laboratory records	707,720	52 (PKU)	52:707,720[b]	7.3	Ia

[a] PKU: classical and atypical combined.
[b] Incidence data for different ethnic groups [cases per 100,000 (95% CI)]: north-west European; 1:12,611 (1: 9512 to 1:17,104), Pakistani 1:14,452 (1:4001 to 1:119,337).

for 6 months significantly decreased plasma triglycerides and significantly increased n-3 LCPUFA. In contrast, the blackcurrant oil led to no significant changes in the lipid profile from baseline, except for n-6 linolenic acid. No adverse reactions were reported during the study period.

In HPA children, treated from first month of life, Agostoni and colleagues[34] found that supplementation with LCPUFA for 12 months improved the docosahexaenoic acid status by around 100% without adversely affecting the arachidonic acid status. No significant changes were observed in blood lipids [total cholesterol, high-density lipoprotein (HDL) cholesterol and low-density lipoprotein (LDL) cholesterol] at the end of treatment.

Treatment during pregnancy

Data from a longitudinal, prospective cohort study (Maternal PKU Collaborative Study) showed that most women with PKU on a low-phenylalanine diet attained metabolic control after 10 gestational weeks (plasma phenylalanine ≤ 605 mmol/l). Scores for the maternal PKU offspring, using the McCarthy General Cognitive Index, decreased as weeks to metabolic control increased ($r = -0.58$, $p < 0.001$). For example, offspring of women with PKU who had metabolic control before pregnancy had a mean ±SD score of 99 ± 13, in contrast to a score of 71 ± 19 for maternal metabolic control achieved after more than 20 weeks. The percentage of children attaining scores 1 (≤ 86) and 2 SD (≤ 72) below the mean increased as metabolic control decreased. For example, 47% of offspring whose mothers did not have metabolic control by 20 weeks of gestation had a General Cognitive Index Score 2 SD below the norm, in contrast to 6% for having metabolic control before pregnancy. Overall, 30% of children born to mothers with PKU had social and behavioural problems.

Discussion

The evidence for the incidence of PKU in the UK is limited to one geographical region of England. Retrospective data suggested that the incidence of PKU (classical and atypical combined) in the West Midlands NHS region was 7.3 cases per 100,000 births between 1981 and 1991.[30] The gene frequency of PKU was much lower among Pakistanis than among north-west Europeans, but the disease incidence was similar in both groups. The effect of consanguineous marriages among Pakistanis may have increased the disease frequency to that seen among Europeans.[30] There are wide variations in the incidence of PKU in the British Isles. Data in the Pollitt report[1] suggest that the average incidence of PKU (classical and atypical combined) in the UK is 11.0 cases per 100,000 live births (England, 10.2 cases per 100,000; Scotland, 12.7 cases per 100,000; and Wales, 9.9 cases per 100,000), whereas in Ireland, the incidence of PKU is much higher at 22.0 cases per 100,000 births. Milder forms of the disease (phenylalanine < 900 μmol/l) accounted for 18% of the total cases in the UK.

The evidence for the effectiveness of specific dietary interventions on blood phenylalanine concentrations,[31] IQ, neuropsychological performance, vitamin B_{12} and folate status for patients with PKU is limited or mainly derived from only one study. The majority of the studies are of poor quality and include small sample sizes; therefore, no firm conclusions can be drawn about the effectiveness of certain dietary interventions for patients with PKU. Moreover, there is a lack of clear evidence about the exact level of phenylalanine restriction, or when, if ever, the restricted diet should be relaxed. Despite these shortcomings, blood phenylalanine concentrations were significantly reduced and vitamin B_{12} levels[36] were higher in subjects with PKU following a phenylalanine-restricted diet compared with those on a less restricted or relaxed diet. The IQ was significantly higher in subjects who continued on the phenylalanine-restricted diet than in those who terminated the diet. Delaying dietary termination until 10 years of age was sufficient to prevent a substantial reduction in cognitive and motor ability; however, there were indications that termination of dietary control at the age of 10 appeared to be associated with deficits in frontal-executive functioning.[35] Other published studies[38] have also found that while dietary termination at 10 years of age or later does not lead to marked deficits in IQ, it does lead to neurological abnormalities.

The subject of dietary management of PKU in adolescence is controversial, and there are differences of opinion among specialists. Rather than just two options, indefinite maintenance of strict dietary control and termination of the special diet at 10 years of age, many specialists suggest partial relaxation of dietary control after the age of 10, with ongoing monitoring of phenylalanine levels and psychological status (Grosse S: personal communication; 2002).

Owing to the limited number of studies and small sample sizes there was insufficient evidence to evaluate the relative effectiveness of tyrosine or

fatty acid supplementation to the diet of people with PKU.

The evidence for the effectiveness of dietary interventions before or during pregnancy is solely based on one longitudinal, prospective cohort study. The authors of this study reported that 30% of children born to mothers with PKU had social and behavioural problems. However, among babies born to mothers with PKU who do not maintain dietary control during pregnancy, at least 90% have one or more severe outcomes, including microcephaly, mental retardation and congenital heart defects (Grosse S: personal communication; 2002). Despite the limited study quality, this study showed that delayed development in offspring of women with PKU is associated with a lack of maternal metabolic control before or early in pregnancy. Treatment at any time during pregnancy may reduce the severity of developmental delay.[37] Therefore, it is extremely important that maternal PKU is managed throughout pregnancy, including during the first trimester.

Conclusion

The conclusions that have been reached by this review for PKU are very similar to those reported by Pollitt and colleagues[1] and Seymour and colleagues.[2] Despite the limited evidence, the average incidence of PKU (classical and atypical combined) in the UK is 11.0 cases per 100,000 live births. Early dietary interventions, including dietary treatment before or during pregnancy, are effective in reducing the severity of developmental delay. Neonatal screening using tandem MS has been shown to be suitable for the reliable detection of PKU.

The authors of the Pollitt report[1] concluded that the UK screening programme for PKU has largely achieved the expected objectives and is cost-effective. Early intervention prevents mental disability, effective therapy is available and a suitable screening test exists. Severe mental retardation due to PKU has all but disappeared in the screened population. Economic evaluations support the view that the replacement of existing screening methods for PKU with tandem MS should only be considered if laboratories plan to extend their neonatal screening programme.

Similar conclusions were reached by the Seymour report.[2] These authors concluded that universal neonatal screening for PKU is worthwhile and should be continued. The estimated incidence of PKU (phenylalanine > 900 μmol/l) in the UK was reported as 1:12,000 live births and the review of the available literature confirms that PKU causes irreversible damage if not treated early, and one in which early treatment, maintained through life at least, results in almost all children developing normally. Cost–benefit analyses show that screening for PKU by itself justifies the collection and testing of neonatal blood spots. Tandem MS has been demonstrated to be robust (accurate, sensitive, lack of false positives) and suitable for the reliable detection of PKU. However, there is insufficient economic evidence to support a change from current methodology to tandem MS-based screening solely for PKU.

Tyrosinaemia type I

Background

Tyrosinaemia type I results from a deficiency of fumarylacetoacetase with resulting accumulation of fumarylacetoacetate and its metabolic precursor maleyl-acetoacetate and inhibition of a variety of other enzyme systems. The characteristics of this disorder are thought to relate to an accumulation of these toxic metabolites.[1,39] Fumarylacetoacetase occurs mainly in the liver and kidneys; therefore, patients with tyrosinaemia type I present with liver failure in infancy (acute form) or show a more protracted course resulting in hepatocellular carcinoma (subacute and chronic forms).

The majority of children with tyrosinaemia type I present with the acute form. Symptoms such as vomiting, diarrhoea, lethargy and failure to thrive may appear in the first months of life. There are also signs of liver disease, with hypoproteinaemia, hyperbilirubinaemia, defective coagulation capacity and hypoglycaemia. Serum tyrosine is elevated and large amounts of tyrosine metabolites are excreted in the urine. Neurological crises similar to those seen in acute porphyria can occur and may be the presenting feature. The condition is progressive and death from liver failure will occur in the first year of life in untreated patients.

The chronic form is characterised by liver disease, renal tubular dysfunction and hypophsophataemia with vitamin D-resistant rickets. Children characteristically present with rickets and/or hepatomegaly during infancy or at school age. Hepatic cirrhosis, hepatoma and renal failure may develop chronically and patients usually die within the first decade of life if left untreated.

Tyrosinaemia types II and III are rare and not covered in this review.

Quantity and quality of research available

Quantity of research available

The yield of literature from the systematic searches generated 84 additional publications on the incidence, prevalence and effectiveness of treatment outcomes for tyrosinaemia type I. Full copies of 17 papers were retrieved for further evaluation.

Of these 17 papers, two studies were judged to be relevant to the update review. One of these studies[30] provided data on the incidence of tyrosinaemia type I within the UK, and the other reported information on the effectiveness of treatment outcomes for tyrosinaemia type I.[40]

Fifteen studies were excluded from the update review (details of references are provided in Appendix 16), for the following reasons:

- publication type: review (n = 3)
- study design: screening (n = 3)
- outcome measures: no treatment outcomes or treatment outcomes in fewer than two patients (n = 4), incidence or prevalence data from non-UK countries (n = 3)
- other reasons: data included in existing systematic reviews (n = 1), poor-quality evidence: treatment duration less than 2 days (n = 1).

Quality of included studies

Incidence of tyrosinaemia type I (Appendix 11)

Only one study reported data on the incidence of tyrosinaemia type I in the UK.[30] This retrospective cohort study evaluated all neonatal infants (n = 707,720) born in the West Midlands NHS region (covering the areas of the West Midlands, Hereford, Worcester, Shropshire, Staffordshire and Warwickshire) during the period 1981–91 for inborn errors of metabolism within different ethnic groups (north-west European, Pakistani, Indian, Afro-Caribbean, other ethnic groups and mixed race). Neonatal samples for screening were obtained 6–10 days after birth (additional details are reported in 'Incidence of PKU', p. 27).

The strength and quality of the evidence for the expected incidence of tyrosinaemia type I in the UK were grade Ia.[30]

Effectiveness of treatments for tyrosinaemia type I

Orthotopic liver transplantation in tyrosinaemia type I (Appendix 17) Only one study was identified that evaluated the effectiveness of orthotopic liver transplantation in patients with tyrosinaemia type I.[40] In this study the authors conducted a retrospective analysis of clinical and biochemical data of patients presenting with tyrosinaemia type I between 1989 and 1997.

During this period, eight children (five males and three females) underwent orthotopic liver transplantation (median age at orthotopic liver transplantation was 64 months, range 5–127 months). Six of these patients developed tyrosinaemia type I between 1989 and 1992, before the availability of 2-(2-nitro-4-trifluoromethylbenzoyl)-1,3-cyclohexanedione (NTBC). The indications for orthotopic liver transplantation were risk of hepatocellular carcinoma with evidence of hepatic dysplasia, chronic liver failure and/or raised α-fetoprotein levels. In contrast, two patients were diagnosed with tyrosinaemia type I between 1992 and 1997, who started NTBC treatment in addition to diet therapy. The indications were non-response to NTBC in one child and development of hepatic dysplasia associated with poor quality of life in the second patient. Of the eight children transplanted, four received a whole liver graft, three received reduction hepatectomy and one had a split liver. During the follow-up period after transplantation (median 6.7 years, range 1–7 years), two children died. Causes of death were primary non-function and chronic rejection.

The strength and quality of the evidence for the effectiveness of orthotopic liver transplantation in patients for tyrosinaemia type I were grade III.[40]

Summary of results

Incidence of tyrosinaemia type I in the UK

Study details and results are summarised in *Table 7*.

The overall incidence of tyrosinaemia type I in the West Midlands NHS region was 1.8 cases per 100,000 births between 1981 and 1991.[30] The incidence of tyrosinaemia type I was significantly higher in the Pakistani population (1:2628, 95% CI 1:1468 to 1:5263) than in the north-west European subjects (1:302,665, 95% CI 1:83,786 to $1{:}2.5 \times 10^6$, $p < 0.001$). These authors also observed that the gene frequency for tyrosinaemia type I was significantly higher in the Pakistani population (1:144, 95% CI 1:87 to 1:271) than in north-west European children (1:550, 95% CI 1:289 to 1:1581, $p < 0.05$). These findings have been attributed to the effect of consanguinity, which increases both anticipated disease and gene frequency rates; alternatively a common

TABLE 7 *Incidence of tyrosinaemia type I in the UK*

Study	Location	Study design	Period of study	Patient selection	Age at sampling	Confirmation of disease	Population screened	Cases diagnosed	Cumulative incidence	Cases per 100,000	Quality of evidence (grade)
Hutchesson *et al.*[30]	Neonatal Screening Service Programme, West Midlands, UK	Retrospective cohort study	April 1981 to April 1991	All newborns in West Midlands NHS region	NR	Yes, laboratory records	707,720	13	13:707,720[a]	1.8	Ia

[a] Incidence data for different ethnic groups [cases per 100,000 (95%CI)]: north-west European, 1:302,665 (1:83,786 to 1:2.5 × 10^6); Pakistani, 1:2628 (1:1468 to 1:5263) ($p < 0.001$). NR, not reported.

causative mutation may be present in the Pakistani population.[30]

Effectiveness of treatment for tyrosinaemia type I

Orthotopic liver transplantation in tyrosinaemia type I

Results from a retrospective analysis of clinical and biochemical data[40] showed that plasma tyrosine and raised α-fetoprotein levels returned to normal in all patients who underwent orthotopic liver transplantation; however, urinary succinylacetone was reduced but persisted in small amounts (median 7.7 μmol/mmol creatinine). Before transplantation the glomerular filtration rate (using height:creatinine ratio) was normal in five out of six survivors. After transplantation three of the six survivors developed renal dysfunction with a fall in the glomerular filtration rate, which stabilised following a reduction in ciclosporin A or a change to mycophenolate mofetil, a powerful immunosuppressive that has no nephrotoxicity. This suggested that the main cause of renal dysfunction was nephrotoxicity due to ciclosporin A.[40]

The clinical outcomes of hypertrophic cardiomyopathy were resolved within 1 year in all patients in whom it occurred ($n = 3$). Before transplantation, one patient did not respond to dietary therapy or NTBC for hypoglycaemia; however, this was resolved following transplantation.

Following orthotopic liver transplantation, two patients died, one from primary non-function and one from chronic rejection. The 1-year actuarial survival rate was 88% and the 5-year actuarial rate was 73%. The quality of life after orthotopic liver transplantation in survivors was reportedly good. All survivors were in mainstream schools and had a reduction in hospital visits (median two per year), venepuncture (median five blood samples per year), medication (median two per day) and dietary freedom, in contrast to patients receiving NTBC therapy who had more hospital visits (median four per year), frequent venepuncture (median 35 blood samples per year), extra medication (median five per day) and a restricted diet.

Discussion

The evidence for tyrosinaemia type I is limited to one geographical region in England. Retrospective data suggest that the incidence of tyrosinaemia type I was 1.8 cases per 100,000 births in the West Midlands NHS region, during the period 1981–91.[30] Moreover, the incidence of tyrosinaemia type I was significantly higher in the Pakistani population (1:2628, 95% CI 1:1468 to 1:5263) than in the north-west European children (1:302,665, 95% CI 1:83,786 to $1{:}2.5 \times 10^6$, $p < 0.001$). Overall, these figures may be an underestimate because the authors reported that they were unable to identify those who died before collection of neonatal sample and children with inborn errors may have died without recognition of the underlying diagnosis, particularly with tyrosinaemia, which can present with neonatal death. However, the authors were confident that many cases of tyrosinaemia type I had not been overlooked.[30]

The evidence for the effectiveness of orthotopic liver transplantation in patients with tyrosinaemia type I was limited to one study. Despite the poor study quality and small sample size, the results showed that orthotopic liver transplantation appeared to be an effective treatment for tyrosinaemia type I, resulting in clinical and biochemical improvements. The quality of life was also enhanced and gave rise to an unrestricted diet, fewer hospital visits and fewer blood samples per annum, in contrast to patients receiving NTBC therapy who had frequent venepuncture and hospital visits and a restricted diet.

Before the introduction of NTBC, the only treatment was orthotopic liver transplantation, but this has an appreciable mortality and morbidity. According to one expert the introduction of NTBC has completely changed this. If NTBC is started early all metabolic changes are reversed and the long-term outlook is excellent. However, if it is started late then liver transplantation may still be necessary because of liver damage and the risk of malignancy (Leonard J: personal communication; 2002).

Most programmes for tyrosinaemia type I are based on the detection of raised tyrosine levels using chromatographic methods or the bacterial inhibition assay method. Tyrosine measurement in dried blood-spot specimens lacks specificity[41–43] because various pathological conditions, including other disorders in tyrosine catabolism[41,44] and benign transient hypertyrosinaemia of the newborn,[41,43,44] can lead to increased tyrosine. In Quebec, Canada, where the prevalence of tyrosinaemia type I is high, detection of succinylacetone is used, as its measurement is reliable for the early diagnosis of tyrosinaemia type I; however, it is too laborious for general neonatal screening.[41]

Several newborn screening programmes outside the UK screen for tyrosinaemia type I using tandem MS. Difficulties in the detection and differentiation of this disorder have been reported by Rashed,[17] Zytkovicz[7] and Muenzer and co-workers.[8] In addition, two false-negative results have been identified so far for tyrosinaemia type I. One case has been identified in New South Wales, Australia,[5] and one case in Bavaria, Germany (Pollitt R: personal communication; 2002). These observations suggest that this disorder is not reliably detected using tandem MS. For optimal sensitivity, Schulze and colleagues[41] recommended a two-tier strategy in neonatal screening of tyrosinaemia type I. Tandem MS was suggested as a first line screening strategy for detecting increased tyrosine concentrations in dried blood spots with cut-off points for tyrosine set between the 90th and 95th percentiles because neonates with hypertyrosinaemia may present with only mildly increased tyrosine concentrations. The second test, an enzyme assay for succinylacetone, was recommended as a confirmatory test for increased tyrosine levels.[41]

Conclusion

The conclusions that have been reached by this review for tyrosinaemia type I are similar to those reported by Pollitt and colleagues[1] and Seymour and colleagues.[2] The expected incidence of tyrosinaemia type I is 1.8 cases per 100,000 births and may be appreciably higher in ethnic groups such as the Pakistani population than in north-west European populations. Effective treatments for tyrosinaemia type I included orthotopic liver transplantations, which improved clinical, and biochemical outcomes and enhanced the quality of life in contrast to NTBC therapy. Evidence regarding the sensitivity and specificity of neonatal screening for tyrosinaemia type I using tandem MS is limited.

The authors of the Pollitt report[1] concluded that the expected incidence of tyrosinaemia type I in the UK ranged from one to five (in some places) cases per 100,000 births. Various types of treatment are available, such as dietary restriction of tyrosine, phenylalanine and methionine, liver transplantation and NTBC treatment. Whichever treatment route is taken, early diagnosis through neonatal screening followed by treatment with NTBC will preserve hepatic and renal function and prevent porphyria-like attacks. However, it is unclear whether NTBC treatment prevents or merely delays the onset of malignancy and whether there are any long-term adverse effects of the drug. Data on the sensitivity and specificity of neonatal screening for tyrosinaemia type I using tandem MS are limited; however, these authors reported that tyrosine can be accurately measured using tandem MS and is suitable as a primary screen.

Seymour and co-workers[2] found that the UK incidence of tyrosinaemia type I was 1:105,000. The incidence of this disorder appeared to be higher in certain parts of the UK owing to the practice of consanguineous marriages or an association with a specific genetic pool. These authors reported that tyrosinaemia is associated with significant morbidity and mortality; however, there are uncertainties regarding the effectiveness of treatments (dietary therapy, liver transplantation and NTBC therapy) and the period before onset during which the intervention improves outcome. Evidence is limited regarding the sensitivity and specificity of neonatal screening for tyrosinaemia type I using tandem MS; however, these authors reported that this method can distinguish between hypertyrosinaemia and benign neonatal tyrosinaemia, which is often associated with hyperphenylalaninaemia. These authors concluded that there is no clear indication for neonatal screening for tyrosinaemia type I in the UK.

Homocystinuria (cystathionine β-synthase deficiency)

Background

Homocystinuria, or cystathionine β-synthase (CβS) deficiency, leads to the accumulation and excretion of homocysteine, homocystine, cysteine–homocysteine mixed disulfide and a number of other sulfur-containing amino acids. Patients appear normal at birth, but a range of clinical and pathological abnormalities involving the eye, skeletal, vascular and central nervous systems develops progressively.[1,39]

A frequent, severe and characteristic feature of this condition is ectopia lentis: the dislocation of the ocular lens, myopia and glaucoma. This is detected in most untreated patients in the first decade of life and in nearly all by the end of the fourth decade. Abnormality of the central nervous system can lead to mental retardation that presents as developmental delay in the first and second years of life. This may be associated with behavioural difficulties and/or convulsions. Osteoporosis is common and this often leads to scoliosis, as well as a tendency towards pathological fractures and vertebral collapse.

Excessive lengthening of the arms, legs and fingers may occur in later childhood, giving a physical appearance similar to that of Marfan syndrome. There is a marked tendency to arterial and venous thrombosis and serious complications can develop in relation to this; for example, optic atrophy secondary to occlusion of the optic artery, hemiparesis, cor pulmonale secondary to pulmonary artery occlusion and seizures or other neurological problems due to cerebral thrombosis. Thrombosis is a major cause of death.

Clinical signs of this serious disease do not become evident until childhood, by which time the complications are largely irreversible. Patients are classified according to whether their biochemical abnormalities are responsive to treatment with pyridoxine (vitamin B_6). Approximately half of all cases respond to this treatment and consequently they have a milder or more slowly developing disease than the unresponsive types.

Several other rare metabolic disorders can also cause homocystinuria,[39,45] but these are not considered in this review.

Quantity and quality of research available

Quantity of research available

The systematic searches identified 194 additional publications on the incidence, prevalence and effectiveness of treatment outcomes for homocystinuria (CβS deficiency). Full copies of 33 papers were retrieved for further evaluation.

Of these 33 papers, five studies were judged to be relevant to the update review. None of these studies provided data on the incidence of homocystinuria (CβS deficiency) in the UK. Based on the best available evidence approach, five studies reported information on the effectiveness of treatment outcomes for homocystinuria.[46–50]

Twenty-eight studies were excluded from the update review (details of references are provided in Appendix 18), for the following reasons:

- publication type: review (n = 3), non-systematic review (n = 1)
- study design: screening (n = 8)
- outcome measures: no incidence or prevalence data (n = 3), no treatment outcomes or treatment outcomes in fewer than two patients (n = 4), incidence or prevalence data from non-UK countries (n = 3)
- other reasons: data included in existing systematic reviews (n = 1), data reported elsewhere (n = 1), not CβS deficiency (n = 3), poor quality evidence: non-availability of synthetic methionine-free, cystine-supplemented mixtures (n = 1).

Quality of included studies

Incidence of CβS deficiency

No additional evidence was available on the incidence of homocystinuria within the UK.

Effectiveness of treatments for CβS deficiency

Dietary interventions for CβS deficiency (Appendix 19)

Five studies were identified that evaluated the effectiveness of dietary interventions for homocystinuria due to CβS deficiency.[46–50]

In the study by Yap and Naughten,[48] the authors conducted a retrospective analysis of clinical and biochemical data of 25 diagnosed cases of homocystinuria due to CβS deficiency. Diagnosis was based on clinical presentation and elevated levels of blood methionine and free homocysteine with low cystine. Patients with homocystinuria consisted of 12 males and 13 females aged between 2.5 and 23.4 years, and all were detected in Ireland between 1971 and 1996 either by the newborn screening programme (n = 21, with blood methionine concentrations > 100 μmol/l and pyridoxine non-responsive) or by clinical presentation (n = 4; three of these were breast-fed and one was pyridoxine responsive). All pyridoxine-non-responsive patients (n = 24) commenced a low-methionine, cystine-enhanced diet (two-thirds of total protein intake was derived from synthetic methionine-free, cystine-supplemented mixture and the remaining third from natural methionine-containing food) supplemented with pyridoxine, vitamin B_{12} and folate within 6 weeks of birth (detected by newborn screening between 8 and 42 days), whereas late-detected cases were started on treatment upon presentation and diagnosis. During follow-up one patient died at 8 years of age in a drowning accident.

In a more recent study, Yap and colleagues[49] evaluated the intellectual abilities of 23 pyridoxine-non-responsive individuals with 339 patient-years of treatment using age-appropriate psychometric tests (full-scale IQ, verbal IQ and performance IQ) and compared the results with those of ten unaffected siblings (control).

All patients diagnosed with homocystinuria due to CβS deficiency were detected in Ireland from 1971 either by the national newborn screening programme or by clinical presentation.[48] All

patients were started on oral pyridoxine while on a normal diet to ascertain their clinical responsiveness to pyridoxine. After confirmation of pyridoxine non-responsiveness (indicated by a persistently high or rising plasma methionine and free homocystine), subjects started on dietary management of methionine restriction and a methionine-free cystine supplemented synthetic amino acid mixture. Patients deficient in plasma vitamin B_{12} and folate were given supplements. Betaine was used as an adjunct to treatment only in those patients (late adolescent/young adults) who became poorly compliant to treatment. In the late-detected pyridoxine-non-responsive patients, betaine was started with cofactor (pyridoxine, vitamin B_{12} and folate) supplementation.

In this study, patients were divided into three groups depending on the time of starting treatment. Newborn-screened patients were further divided into two groups depending on their compliance, as defined by their biochemical control (criterion for good compliance was defined as a treatment lifetime plasma-free homocystine median < 11 μmol/l). Nineteen patients were detected through newborn screening. Of these, 13 were compliant with treatment (mean age 14.4 years, range 4.4–24.9 years) and the remaining six, who were poorly compliant, developed complications (mean age 19.9 years, range 13.8–25.5 years). All newborn-screened patients had begun a methionine-restricted, cystine-supplemented diet within 6 weeks of birth. Two patients who were detected late (aged 18.9 and 18.8 years) were started on treatment at the ages 2.4 and 2.9 years, respectively, and two were untreated at the time of assessment (aged 22.4 and 11.7 years at diagnosis). The control group (mean age 19.5 years, range 7.8–32.9 years) consisted of ten unaffected siblings. Two patients, who were detected through newborn screening, were excluded from this study: one died from drowning, and psychometric data from one patient (a postgraduate psychology student) may not have been valid owing to her knowledge, training and refusal to be assessed.

In one study,[46] using a before and after intervention design, the authors evaluated the effect of thiamin (vitamin B_1) supplementation in nine homozygote patients with homocystinuria due to CβS deficiency. All nine patients had good compliance to their homocysteine-lowering treatment, had prolonged elevated fasting homocysteine levels in the blood despite homocysteine-lowering treatment for at least 2 years and were at least 14 years of age. Homozygosity for homocystinuria was confirmed by hypermethioninaemia, severe hyperhomocysteinaemia and almost absent cystathionine activity in cultured skin fibroblasts. Only one of the patients was on a regimen of methionine restriction. In addition to conventional homocysteine-lowering therapy (pyridoxine, vitamin B_{12}, folic acid and/or betaine), six patients received 25 mg of thiamin hydrochloride for 6 weeks three times daily, and three patients twice daily. These three patents received only in total 50 mg thiamin hydrochloride conforming with their conventional homocysteine-lowering treatment, which was also twice daily. All patients were aged between 14 and 34 years (mean age 24.4 ± 6.2 years) with a mean duration of homocysteine-lowering treatment of 9.0 ± 2.4 years (range 5–17 years). The mean thiamin dosage for the patients was 0.90 ± 0.25 mg thiamin supplementation/kg body weight (range 0.48–1.43 mg/kg).

Wilcken and Wilcken[47] presented the vascular disease findings in homocystinuria due to CβS deficiency seen while managing all patients identified with this disorder in the state of New South Wales, Australia. In this retrospective study of clinical and biochemical data, 40 patients were diagnosed on the basis of characteristic clinical features and elevated levels of plasma methionine and free homocyst(e)ine with low free cyst(e)ine. During the follow-up period, ten patients died at the age of 2–30 years. Only two of these ten patients were receiving effective treatment at the time of death. One patient died in an accident unrelated to homocystinuria and one died of a pulmonary embolus; however, it was unclear whether this patient complied with treatment.

All patients had increased urinary homocysteine and were categorised as being either pyridoxine responsive [plasma homocyst(e)ine < 20 μmol/l] or pyridoxine non-responsive. All treated patients received pyridoxine 100–200 mg/day and folic acid 5 mg/day and most had intermittent hydroxocobalamin by injection according to serum vitamin B_{12} status, measured twice yearly. Pyridoxine-non-responsive patients all received, in addition, 6–9 g/day oral betaine given in two divided doses. Diet was not closely monitored, but general advice was given to reduce the intake of foods with a high methionine content. Data from 32 patients in this study, who were followed up with treatment, were included in the study by Yap and colleagues.[50]

Yap and colleagues[50] performed a multicentre observational study to assess the effectiveness of

long-term homocysteine-lowering treatment in reducing vascular risk in 158 patients. Vascular outcomes were analysed and the effectiveness of treatment in reducing vascular risk was evaluated by comparison of actual to predicted number of vascular events, with the use of historical controls from a landmark study of 629 untreated patients with CβS deficiency.

From the five centres in Ireland (Dublin, $n = 27$), Australia (Sydney, $n = 32$), The Netherlands (Nijmegen, $n = 28$) and the UK (Manchester, $n = 30$; London, $n = 41$), 158 patients with CβS deficiency (pyridoxine responsive, $n = 70$; pyridoxine non-responsive, $n = 88$) were treated and followed up for a mean of 17.9 years (pooled data). The mean age was 29.4 years (range 4.5–70 years). Only 3% were younger than 10 years of age, 36% were older than 30 years and 6.3% were older than 50 years.

Pretreatment plasma free homocysteine levels were between 11 and 187 μmol/l (Dublin, Manchester and London) and plasma total free homocysteine levels were between 42 and 266 μmol/l (Sydney and Nijmegen). Three main treatment regimens were used by all centres, with minor modifications (Appendix 19). Initially, therapeutic doses of pyridoxine (vitamin B_6) were given in combination with folate. If the response to vitamin B_6 was inadequate (pyridoxine-responsive patients were those whose total free plasma homocysteine was reduced to < 20 μmol/l with pyridoxine treatment), dietary methionine restriction was attempted. However, vitamin B_6-non-responsive pyridoxine continued to be given to many of the patients owing to reports of its beneficial effects. The third therapeutic option was the use of betaine, mainly in patients who were non-responsive to pyridoxine. Despite the slightly different combinations of treatment regimens used by the five centres, the authors reported that there was no evidence of non-homogeneity in the vascular outcomes of each centre ($p = 0.156$).

The biochemical markers used for monitoring homocysteine and the criteria for assessing the response to pyridoxine differed among centres. The frequencies of biochemical monitoring in Sydney, Nijmegen, Manchester and London were similar (between one and four times per year), whereas in Dublin it was monitored more often (at least eight to ten times per year). The authors reported that some patients included in the study were taking other treatment than that specifically designed to lower homocysteine. Some older patients ($n = 23$) were taking aspirin and one was undergoing lipid therapy. No details regarding compliance with treatment were provided.

The strength and quality of the evidence for the effectiveness of dietary interventions for homocystinuria (CβS deficiency) were grade II-2a,[49] grade II-2b[50] and grade III.[46–48]

Summary of results

Incidence of CβS deficiency

No additional evidence was available on the incidence of homocystinuria (CβS deficiency) in the UK. Owing to the lack of additional evidence, incidence data reported in the existing systematic reviews[1,2] are summarised in *Table 8*.

Effectiveness of treatments for CβS deficiency

Dietary interventions for CβS deficiency

Data over a 25-year period from Ireland[48] found 25 diagnosed cases of CβS deficiency. Of these 25 patients, 24 were non-responsive to pyridoxine. Clinical and biochemical data suggest that no homocystinuria-related complications were found in 18 dietary-treated cases. In these 18 subjects, three developed increasing myopia and all had higher lifetime median free homocysteine levels (range 18–48 μmol/l) compared with the remaining 15 patients, who had lifetime median free homocysteine levels below 11 μmol/l.

Among the three screened non-dietary-compliant cases, two presented with ectopia lentis, one had osteoporosis and two had mental disability. In this group, poor dietary compliance (reflected by diet history and higher plasma free homocysteine levels) for between 2 and 8 years significantly increased the risk of developing complications. In the four patients who were missed by screening, three presented with ectopia lentis after the age of 2 years.

All 25 patients had lifetime median methionine levels ranging from 47 to 134 μmol/l and no patients developed any thromboembolic events (age range 2.5–23.4 years). Four patients, who were non-compliant to dietary regimens ($n = 2$) or were missed by screening and presented with complications after 2 years of age ($n = 2$) had mental disability, whereas the remaining 21 patients achieved age-appropriate education standards.

Yap and colleagues[49] evaluated the intellectual abilities of 23 pyridoxine-non-responsive individuals and found that there were no statistically significant differences between the newborn-screened compliant group and

TABLE 8 *CβS deficiency detected by screening in the British Isles*[a]

Region/country	No. screened	No. detected	Incidence	Cases[b] per 100,000 births
NW Health Region, England	1,257,179	10	1:126,000	0.8
Scotland	1,012,500	1	1:1,000,000	0.1
Northern Ireland	549,219	7	1:78,500[c]	1.3
Republic of Ireland	1,341,272	23	1:58,000	1.7

[a] Original data collected by Mudd *et al.*[51]
[b] Homocystinuria due to CβS deficiency.
[c] Seymour *et al.*[2] reported the incidence as 1:78,000.

unaffected control group in the psychometric parameters assessed, except for a higher full-scale IQ in the patient group ($p = 0.0397$).

The newborn screened, good compliance group ($n = 13$), with a total of 187.7 patient-years of treatment, had a mean full-scale IQ of 105.8 (range 84–120), verbal IQ of 110.3 (range 88–117) and performance IQ of 98.6 (range 78–118). These patients did not have any other recognised homocystinuria-related complications and had lifetime free homocystine medians less than 11 μmol/l. In contrast, the control group ($n = 10$) had a mean full-scale IQ of 102 (range 76–116), verbal IQ of 107 (range 81–123) and performance IQ of 96.6 (range 76–115).

Patients in the newborn screened, poorly compliant group ($n = 6$) had a mean full-scale IQ of 80.8 (range 40–103), verbal IQ of 87.3 (range 46–113) and performance IQ of 75.2 (range 46–87). Correspondingly, the lifetime free homocystine medians were inversely related to full-scale IQ. The patients in this group received a total of 118.9 patient-years of treatment.

The two late-detected patients had a full-scale IQ of 80 and 102, verbal IQ of 75 and 107, performance IQ of 92 and 94, and lifetime free homocystine medians of 4.5 and 8.5 μmol/l, respectively. The two untreated patients had full-scale IQ of 52 and 53, verbal IQ of 58 and 61, and performance IQ of 52 and 52, respectively, before starting treatment.

A study that evaluated the effect of thiamin (vitamin B_1) supplementation in nine homozygotes patients with homocystinuria due to CβS deficiency[46] found that the thiamin plasma concentration increased from 123 ± 30 nmol ($n = 7$) before thiamin treatment to 207 ± 38 ($n = 7$) after 6 weeks of treatment. The fasting blood methionine concentration significantly decreased from 194 ± 127 to 128 ± 94 μmol/l after thiamin treatment ($n = 9$; $p < 0.06$). The mean fasting plasma homocysteine concentrations did not differ significantly before and after thiamin treatment (91 ± 25 μmol/l to 83 ± 22 μmol/l, respectively), however, of the nine patients, only one showed a substantial reduction in the homocysteine concentration simultaneously with a major decrease in the methionine level. In all patients, the level of the serum transamination metabolites remained unchanged..

Results from a long-term follow-up of patients with homocystinuria due to CβS deficiency[47] found that of the 32 patients (mean age 30 years, range 9–66 years) receiving effective treatment (pyridoxine, folic acid and hydroxocobalamin), 17 were pyridoxine responsive and all maintained plasma total free homocyst(e)ine levels < 20 μmol/l over an average treatment period of 16.6 years. In this group, there were two vascular deaths, one fatal pulmonary embolus and one myocardial infarction, whereas without treatment 21 deaths would have been expected ($\chi^2 = 14.22$, $p = 0.0001$, relative risk 0.09, 95% CI 0.02 to 0.38). In contrast, 15 patients who were non-responsive to pyridoxine also received 6–9 g of betamine of betaine daily, which resulted in a substantial decline (mean 74 ± 14%) in plasma total free homocyst(e)ine levels, persisting during an average (postbetaine) treatment period of 11 years (mean ± SD levels: 33 ± 17 μmol/l). In this group, there were no vascular events during 258 patient years of treatment ($p < 0.005$ versus expected untreated). Nineteen subjects had a total of 19 major and 15 minor operations requiring anaesthetic, and there were three successful pregnancies, one in a patient receiving betaine. There were no thromboembolic complications.

A multicentre observational study[50] that investigated the effect of long-term homocysteine-lowering treatment on cardiovascular risk in 158 patients had a total of 2821.6 patient-years of treatment. There were 1243.8 patient-years of

treatment in the vitamin B_6 responders, and a further 1577.8 patient-years of treatment in the vitamin B_6 non-responders.

Without treatment, 112 vascular events would have been expected (one vascular event per 25 years was expected in 2821.6 patient-years of treatment). Instead, only 17 vascular events (pulmonary embolism, $n = 3$; myocardial infarction, $n = 2$; abdominal aortic aneurysm, $n = 2$; transient ischaemic attack, $n = 1$; sagittal sinus thrombosis, $n = 1$; deep vein thrombosis, $n = 5$; and cerebrovascular accident, $n = 3$) occurred in 12 patients who were undergoing treatment [relative risk (RR) 0.09, 95% CI 0.036 to 0.228, $p < 0.0001$]. Of the seventeen vascular events, 12 occurred in eight vitamin B_6 responders at a mean age of 51.6 years (range 25–67 years) and five vascular events in four vitamin B_6-non-responders at a younger mean age of 20.6 years (range 18–24 years). There were five vascular deaths during the treatment period, two in the vitamin B_6 responders (pulmonary embolism, myocardial infarction) and three among vitamin B_6 non-responders (pulmonary embolism, sagittal sinus thrombosis). In the total of 825 patient-years of betaine treatment (the longest period of betaine treatment was 17 years) there were no reports of significant side-effects. Plasma homocysteine levels were markedly reduced from pretreatment levels, but were several times higher than the mean for the respective normal population of each centre, despite the aim of achieving normal levels.

Discussion

Despite the lack of additional evidence on the incidence of homocystinuria (CβS deficiency) in the UK, data reported in the two existing systematic reviews[1,2] suggest that the incidence of CβS deficiency was 0.8 cases per 100,000 births (1:126,000) in England, 0.1 cases per 100,000 births (1:1,000,000) in Scotland and 1.3 cases per 100,000 births (1:78,500) in Northern Ireland. These authors believed that these estimates of expected incidence hide much regional variation[2] within the UK and there may be underestimates of cases of homocystinuria since the pyridoxine-responsive disease is not often detected using current methods of neonatal screening.[1,2]

The evidence for the effectiveness of dietary interventions in patients with CβS deficiency was limited to a few studies. Despite the limited study quality, small sample sizes (in some studies) and varying length and types of treatment, the result showed that homocystinuria due to CβS deficiency was a potentially treatable disorder over a long period. Newborn screening, early treatment, good dietary compliance and the maintenance of a lifetime median plasma free homocysteine less than 11 μmol/l appeared to protect against the overt recognised complications of untreated homocystinuria[48] and prevent mental retardation in pyridoxine-non-responsive CβS deficient patients.[49] The treatment regimens designed to lower plasma homocysteine significantly reduced cardiovascular risk in CβS-deficient patients, despite imperfect biochemical control;[50] however, the authors of this study also acknowledged that it was not clear whether this results entirely from lowering the extremely high pretreatment levels of homocysteine or from some other aspect of the treatment. Furthermore, the additional effect of betaine therapy in pyridoxine-non-responsive patients is substantial and produces sustained reduction of elevated plasma homocyst(e)ine, thereby contributing to reducing cardiovascular risk.[47]

Newborn screening for homocystinuria is conducted in the USA, Japan and many European countries.[52] Most programmes for homocystinuria are based on the detection of high methionine levels (hypermethioninaemia) using the Guthrie bacterial inhibition assay. This method mostly appears to detect those patients not responding to pyridoxine, and it is increasingly becoming apparent that many pyridoxine-responsive cases, who may account for 50% of affected infants, are being missed by the screening programmes.[1,51,53] Reducing the cut-off level for methionine from 134 to 67 μmol/l has substantially increased the frequency of identified infants in New England, USA; however, this was associated with a higher false-positive rate, which increased from 0.006 to 0.03%.[54] Tandem MS, which has a greater sensitivity for methionine than the bacterial inhibition assay,[55] is unlikely to eliminate false-negative results, because the blood methionine concentration may not be elevated during the first few days of life.[1] Since 1998, newborn screening using tandem MS has been used routinely in the New South Wales, Australia. Based on prospective and retrospective results, Wilcken and colleagues[5] concluded that homocystinuria (as well as mild MSUD, tyrosinaemia type I and probably some but not all fatty acid oxidation disorders) could not be reliably detected using tandem MS; however, there was no justification in not seeking a disorder simply because it cannot be detected reliably on every occasion.[5]

Conclusion

The conclusions that have been reached by this review for homocystinuria due to CβS deficiency are similar to those reported by Pollitt and colleagues.[1]

Data reported in the two existing systematic reviews[1,2] suggest that the incidence of CβS deficiency is 0.8 cases per 100,000 births (1:126,000) in England, 0.1 cases per 100,000 births (1:1,000,000) in Scotland and 1.3 cases per 100,000 births (1:78,500) in Northern Ireland. These estimates of expected incidence hide many regional variations within the UK. Despite the limited quality, small sample sizes (in some studies) and varying lengths of treatment, homocystinuria due to CβS deficiency is a potentially treatable disorder (dietary, pyridoxine and betaine therapy) over a long period. Newborn screening, early treatment and good compliance are highly effective in preventing long-term complications; however, the response to treatment varies depending on whether patients are pyridoxine responsive or pyridoxine non-responsive. Evidence for the detection (sensitivity and specificity) of homocystinuria (by hypermethioninaemia) using tandem MS is limited.

The authors of the Pollitt report[1] reported that CβS deficiency is a serious disease and is clinically silent until childhood, by which time the complications are largely irreversible. Treatment (dietary restriction of methionine with cystine supplementation, pyridoxine and betaine supplementation) will either prevent or greatly delay the progression of the disease; however, the response to treatment varies with both the severity of the biochemical lesion (particularly whether pyridoxine responsive or non-responsive) and the age at which treatment is started. Data on the sensitivity and specificity of neonatal screening for homocystinuria due to CβS deficiency using tandem MS are limited; however, these authors concluded that persistent hypermethioninaemia in the absence of homocystinuria or severe liver diseases is very rare; thus, the false-positive rate with tandem MS is likely to be low.

Seymour and colleagues[2] reported that homocystinuria due to CβS deficiency causes significant morbidity and mortality (impaired vision, osteoporosis, neurological dysfunction and thromboembolism) and there are uncertainties regarding the effectiveness of treatments (dietary restriction and pyridoxine supplementation) for homocystinuria. Evidence is limited regarding the sensitivity and specificity of neonatal screening for homocystinuria using tandem MS; however, these authors reported that this method offers a sensitivity of 4 μmol/l for methionine, as well as a demonstrated ability to reduce the rate of false positives compared with the bacterial inhibition assay method and to reduce significantly false negatives. These authors concluded that there is no justification for including neonatal screening specifically for homocystinuria in the UK screening programme.

MSUD

Background

MSUD is caused by an inherited deficient activity of branched-chain 2-keto acid dehydrogenase affecting the metabolism of all three branched-chain oxo and amino acids. Four major variants of MSUD have been identified: severe (classical), intermediate, intermittent and thiamin responsive.

In the classical presentation of this disease there is little residual enzyme activity, and branched-chain amino acids and keto acids accumulate in the body fluids, resulting in the sweet, caramel-like or 'maple syrup' smell apparent in the urine of some patients. Patients are apparently normal at birth, with symptoms associated with ketoacidosis such as poor feeding, vomiting and lethargy occurring in the first week of life. The very low levels of enzyme activity associated with this severe form of the condition lead to rapid progression of the disease, with neurological symptoms and weight loss at 7–10 days followed by increasing lethargy, convulsions and irregular respiration. Severe ketoacidosis may lead to coma and death, or survival with neurological damage including mental retardation, blindness and cerebral palsy.

The higher residual enzyme activity in the intermediate and intermittent forms of the disease generally delays the onset of symptoms, which are less severe and may only present (in the intermittent form) at times of catabolic stress such as infection. A thiamin-responsive form has been described that is clinically less severe than the other forms, but this is extremely rare.

Quantity and quality of research available

Quantity of research available

The yield of literature from the systematic searches generated 59 additional publications on the incidence, prevalence and effectiveness of treatment outcomes for MSUD. Full copies of 22 papers were retrieved for further evaluation.

Of these 22 papers, none was judged to be relevant to the update review. These 22 studies were excluded from the update review (details of references are provided in Appendix 20), for the following reasons:

- publication type: review (n = 4)
- study design: screening (n = 10)
- outcome measures: no incidence or prevalence data (n = 1), treatment outcomes in fewer than two patients (n = 1), incidence or prevalence data from non-UK countries (n = 5)
- other reasons: data included in existing systematic reviews (n = 1).

Quality of included studies

Incidence of MSUD

No additional evidence was available on the incidence of MSUD in the UK.

Effectiveness of treatments for MSUD

No additional evidence was available on the effectiveness of treatments for MSUD.

Summary of results

Incidence of MSUD

No additional evidence was available on the incidence of MSUD in the UK. Owing to the lack of additional evidence, incidence data from the existing systematic reviews[1,2] have been reported.

Pollitt and colleagues[1] found that the expected incidence of MSUD in the UK was 0.5 cases per 100,000 births. These data were extrapolated from whole population screening or comprehensive national surveys of clinically detected cases for non-UK (but broadly similar) populations. Seymour and colleagues[2] reported that in Scotland, no cases were detected in 467,448 newborn infants screened.

Effectiveness of treatments for MSUD

No additional evidence was available on the effectiveness of treatments for MSUD. Despite this, data reported in the two existing systematic reviews suggest that various treatments are available for this disorder. Dietary therapy is for life[2] and requires a reduction in branched-chain amino acid intake;[1] however, it is highly restrictive, is difficult to maintain and does not completely prevent recurrent metabolic crises, which remain potentially fatal.[2] Management of the severe forms of MSUD involves peritoneal dialysis, haemodialysis or haemofiltration to reduce metabolite levels.[1] Additional metabolic irregularities due to infections in later infancy and childhood necessitate emergency support with intravenous fluids (glucose with concomitant insulin or lipids).[1]

Discussion

Despite the lack of additional evidence on the incidence of MSUD in the UK, data reported in the two existing systematic reviews[1,2] suggest that the incidence of the disease was relatively low in the UK compared with elsewhere. Pollitt and colleagues[1] reported that the expected incidence of MSUD in the UK was 0.5 cases per 100,000 births; however, these authors believed that in certain areas these estimates of expected incidence would be higher in immigrant populations with a high rate of consanguineous marriage.

Conclusion

No additional evidence was available on the UK incidence and effectiveness of treatments for MSUD. Information for this disorder was solely derived from the two existing systematic reviews.[1,2] Evidence regarding the sensitivity and specificity of neonatal screening for MSUD using tandem MS is limited.

Pollitt and colleagues[1] reported that the expected incidence of MSUD in the UK was 0.5 cases per 100,000 births; however, this estimate was expected to be substantially higher in immigrant populations with a high rate of consanguineous marriage. These authors found that patients with the severe form of the disease need to be diagnosed and treated (peritoneal dialysis, haemodialysis or haemofiltration and dietary restriction) in the first week of life to ensure a good outcome. These authors concluded that the timing of sample collection in the UK is not ideal for MSUD screening.

Seymour and colleagues[2] found no cases of MSUD (0:467,448) in a newborn screening programme conducted in Scotland. These authors reported that MSUD is associated with significant morbidity and mortality; however, there are uncertainties regarding the effectiveness of treatments (dietary therapy is for life, is difficult to maintain and does not completely prevent recurrent metabolic crises, which remain potentially fatal) and the period before onset during which the intervention improves outcome. Tandem MS offers a sensitivity of 2 μmol/l for leucine plus isoleucine, reduced false positives and potentially fewer false negatives. These authors concluded that there is no justification for including neonatal screening specifically for MSUD in the UK screening programme, as it does not meet all of the criteria required for the neonatal screening programme.

Urea cycle disorders

Background

The urea cycle is a metabolic pathway, confined to the liver, that leads to the detoxification of

ammonia by the synthesis of arginine and urea. There are five urea cycle disorders, each relating to a defect in one of the enzymes of the urea cycle: carbamoyl phosphate synthase deficiency, ornithine carbamoyltransferase (transcarbamylase) deficiency, argininosuccinate synthase deficiency/citrullinaemia, argininosuccinate lyase deficiency/argininosuccinic aciduria, and arginase deficiency/argininaemia.[1,2]

Any disruption in the synthesis of urea leads to accumulation of the ammonium ion, which is highly toxic; therefore, most of the urea cycle disorders share a similar spectrum of clinical presentation. Hyperammonaemia is thought to be the main damaging factor in the first four disorders and they share many common features. These will be discussed as a group distinct from arginase deficiency, which has a somewhat different presentation.

The disorders carbamoyl phosphate synthase deficiency, ornithine carbamoyltransferase (transcarbamylase) deficiency, argininosuccinate synthase deficiency/citrullinaemia and argininosuccinate lyase deficiency/argininosuccinic aciduria have variable severity with a neonatal acute form that is rapid and usually fatal. They all have a range of milder variants that present as chronic conditions later in infancy and childhood. The acute neonatal presentation of these disorders is seen in full-term infants who present with the effects of hyperammonaemia in the first days of life. Acute presentation involves respiratory distress, poor feeding, lethargy, vomiting, hypotonia, spasticity, convulsions and coma. Pulmonary and gastric haemorrhages can also occur and in untreated cases most patients will die in the neonatal period. Those who survive the first few days of life probably have some residual enzyme activity.

The chronic forms of these urea cycle disorders often present with a history of episodic vomiting, lethargy and irritability, and there can be seizures or even periods of coma. These episodes are often associated with high-protein meals or minor infections. Some of these late-onset or chronic presentations of the disorders remain healthy until later childhood, when they suffer acute illness often associated with an infection. Undiagnosed and untreated chronic cases may prove fatal.

Arginase deficiency is a distinct condition insofar as there are usually only minor symptoms of irritability and mild developmental delay in infancy. However, over the next few years progressive spasticity, tremor, ataxia, choreoathetosis, seizures and severe mental retardation will develop. These symptoms are thought to be related to the toxic effects of arginine rather than hyperammonaemia.

Another rare disorder, hyperornithinaemia, homocitrullinuria, hyperammonaemia (HHH) syndrome, also impedes the urea cycle. The clinical presentation of this disorder appears to be related to hyperammonaemia, and intolerance to protein feeding, vomiting, seizures and developmental delay are common. Severe neonatal cases present with lethargy, hypotonia and seizures that can progress to coma and eventual death.

This review focuses on the following disorders: ornithine carbamoyltransferase (transcarbamylase) deficiency, citrullinaemia, argininosuccinic aciduria, arginase deficiency and hyperornithinaemia, as these have been reported to be detectable by tandem MS.[1]

Quantity and quality of research available

Quantity of research available

The systematic searches identified 147 additional publications on the incidence, prevalence and effectiveness of treatment outcomes for urea cycle disorders. Full copies of 22 papers were retrieved for further evaluation.

Of these 22 papers, two studies were judged to be relevant to the update review. Both publications[56,57] reported information on the effectiveness of treatment outcomes for urea cycle disorders, namely ornithine transcarbamylase deficiency and citrullinaemia. No studies were identified that provided data on the incidence of urea cycle disorders in the UK.

Twenty studies were excluded from the update review (details of references are provided in Appendix 21), for the following reasons:

- patients: not urea cycle disorder (n = 2)
- study design: screening (n = 5)
- outcome measures: mutation, genotype, etc., analysis (n = 5), no treatment outcomes or treatment outcomes in fewer than two patients (n = 5), incidence or prevalence data from non-UK countries (n = 3).

Quality of included studies

Incidence of urea cycle disorders

No additional evidence was available on the incidence of urea cycle disorders in the UK.

Effectiveness of treatments for urea cycle disorders (Appendices 22 and 23)
Two studies were identified that evaluated the effectiveness of dietary interventions for urea cycle disorders, namely ornithine transcarbamylase deficiency and citrullinaemia.

In one study,[56] the authors evaluated the safety and efficacy of sodium phenylbutyrate in long-term therapy of nine patients with ornithine transcarbamylase deficiency using retrospective analysis of clinical and biochemical data from three different European paediatric centres.

Patients with ornithine transcarbamylase deficiency consisted of four males and three females who were aged between 6 days and 14 years at diagnosis and started sodium phenylbutyrate therapy at 2–16 years of age. All patients were diagnosed with ornithine transcarbamylase deficiency by liver enzyme assay and/or DNA mutation analysis. These nine patients had previously been treated with oral sodium benzoate (median dose 248 mg/kg per day, range 106–275 mg/kg per day) and a low-protein diet (median 0.84 g/kg per day), arginine (n = 5) or citrulline (n = 4) supplementation for at least 8 months (median 17 months, range 8–28 months). Sodium benzoate was replaced when phenylbutyrate became available in European countries. Sodium phenylbutyrate was given in three or four divided doses (median 352 mg/kg per day, range 125–484) at 8.9 and 4.9 years of age (median) to males and females, respectively. The median duration and follow-up was 26 months, ranging from 17 to 42 months.

The authors reported that in this retrospective study, it was not possible to obtain accurate data because patients were referred from different centres having distinct nutritional protocols and that the treatment regimen was not sufficient for the increase in protein intake allowed. Furthermore, the authors did not report whether any patients were lost to follow-up or whether patients were compliant with treatment.

Maestri and colleagues[57] monitored the long-term survival and outcomes of 24 patients with neonatal-onset argininosuccinate synthase deficiency (citrullinaemia) using clinical and biochemical data from 18 medical institutions throughout the USA and Canada. Patients in this study were born before 1990 and diagnosed with citrullinaemia within the first month of life. Each neonate was the healthy product of a term pregnancy with normal anthropometric measurements. Fifteen infants were white, two were Hispanic, one was black, five were of unknown or mixed racial background and no details were provided for one patient. Diagnosis of argininosuccinate synthase deficiency was based on elevated plasma ammonium levels (ranging from 266 to 2000 μmol/l), increased plasma citrulline levels (> 1000 μmol/l) and no detectable plasma argininosuccinate.

Specific therapeutic protocols were designed to activate alternative pathways of waste nitrogen excretion. During the 15-year study period, dietary treatment protocols were maintained by protein restriction (range 0.5–2.0 g natural protein/kg body weight per day), whereas pharmacological interventions were modified owing to the availability of new drugs. Protocol I included administration of sodium benzoate (250 mg/kg per day) and arginine (500–700 mg/kg per day) freebase (subjects born before 1984 were maintained on this protocol). Protocol II also included sodium phenylacetate (protocol IIa), first used in 1984, or sodium phenylbutyrate (protocol IIb), first used in 1985 (250 mg/kg per day). Protocol III included high doses (450–600 mg/kg per day) of sodium phenylacetate (protocol IIIa) or phenylbutyrate (protocol IIIb), but excluded sodium benzoate; patients were transferred to this protocol as it became available in 1987. The number of patients treated in each protocol regimen is not explicitly clear.

Participation in this study was voluntary and transfer to a newer protocol was neither strictly enforced nor strongly recommended until evidence accumulated that a newer therapy was effective. The different treatment histories for each patient make it difficult to compare the effectiveness of each protocol. The authors provided no information on age or gender, and nine patients were lost to follow-up (three died during treatment protocol IIa, four died on protocol III regimen and two withdrew from therapy).

The strength and quality of the evidence for the effectiveness of dietary and pharmacological interventions for ornithine transcarbamylase deficiency and citrullinaemia were grade III.[56,57]

Summary of results

Incidence of urea cycle disorders

No additional evidence was available on the incidence of urea cycle disorders in the UK. Owing to the lack of additional evidence, incidence data from the existing systematic reviews[1,2] have been reported.

Pollitt and colleagues[1] reported that there were no reliable data available on the expected incidence of urea cycle disorders within the UK. Therefore, the expected incidence of urea cycle disorders was estimated from the number of cases clinically diagnosed in the UK (compared with similar disorders of established incidence). This was found to be 2.5 cases per 100,000 births. Seymour and colleagues[2] found that in a neonatal screening programme, based on screening of urine using chromatography, no cases of arginosuccinate lyase, arginase or argininosuccinate synthase deficiency were detected in 135,295 infants screened in Wales. These authors reported that the true incidence of urea cycle defects, particularly ornithine carbamoyltransferase deficiency, was unknown.[2]

Effectiveness of treatments for urea cycle disorders

In the long-term management of nine patients with ornithine transcarbamylase with a low-protein diet and sodium phenylbutyrate, Burlina and colleagues[56] found that sodium phenylbutyrate was well tolerated, no adverse effects were detected during the treatment period and there were no hyperammonaemic episodes requiring hospitalisation. The total protein intake increased from 0.84 g/kg per day (range 0.43–1.63 g/kg per day) before starting treatment with phenylbutyrate to 0.95 g/kg per day (range 0.66–1.46 g/kg per day) after 18 months of treatment. Total protein intake correlated significantly with increases in the dose of phenylbutyrate ($p < 0.001$), suggesting that an increase in the phenylbutyrate dose allowed an increase of 1 g in protein intake. In contrast to the decline in plasma ammonia levels, plasma glutamine levels did not show any significant variation during the treatment period. Plasma concentrations of essential amino acids including alanine, asparagines, leucine and isoleucine remained normal. In addition, other routine haematological and biochemical measurements (liver and renal function tests) remained normal during the treatment period, suggesting a lack of toxicity effect. No cognitive evaluations were performed during the treatment, but the authors presumed that the metabolic stability might have prevented neurological deterioration over the period of treatment.

In 24 patients born before 1990 with neonatal citrullinaemia and treated since birth with various therapeutic protocols designed to limit dietary nitrogen, the cumulative survival rate was 87.5% at 5 years and approximately 72% at 10 years of age.[57] Overall, 15 patients survived (three died during treatment protocol IIa, four died on protocol III and two withdrew from therapy) during treatment with high doses of sodium phenylbutyrate. Among these 15 surviving patients, 11 were classified as severely to profoundly mentally retarded (IQ < 55). The remaining four patients had IQ measurements in the low borderline to mentally retarded range (IQ between 50 and 70). All patients had intercurrent hyperammonaemic episodes, the frequency of which decreased with the implementation of protocol IIIb. However, there was wide variability in the number and frequency of episodes in individual patients. On average, the 15 surviving patients had one episode per year (range 2–30 episodes during 5.4–15.5 years of treatment). Overall, patients were growth retarded and had height for weight *Z*-scores within 2 SD of the mean. Laboratory studies of plasma amino acids and of haematopoietic, renal and hepatic function were within normal limits, with the exception of slightly elevated serum aminotransferase values.

Discussion

Despite the lack of additional evidence on the incidence of urea cycle disorders within the UK, information reported in the two existing systematic reviews[1,2] suggests that there are no reliable data available on the expected incidence of urea cycle disorders in the UK and that there are unexplained marked geographical variations in the incidence between populations of potentially similar ethnic origin namely, Australia, Canada and the USA.[2] Within the urea cycle group of diseases, Pollitt and colleagues[1] reported that ornithine carbamoyltransferase deficiency was the most frequently diagnosed disorder, followed by argininosuccinate synthase (citrullinaemia) and argininosuccinate lyase deficiency.

The evidence for the effectiveness of dietary and pharmacological interventions in patients with ornithine transcarbamylase deficiency and citrullinaemia was limited to a few studies. Despite the poor study quality, small sample sizes and different nutritional protocols, the results showed that dietary protein restriction and sodium phenylbutyrate facilitated the clinical management of patients with ornithine transcarbamylase deficiency and enhanced their quality of life.[56] The survival rates for patients with citrullinaemia also improved;[57] however, in this study seven out of 24 patients with citrullinaemia died, a further two were lost to follow-up, and survival for patients with citrullinaemia was accompanied by mental retardation, growth retardation, risk of hyperammonaemic episodes and the need for

lifetime adherence to strict medication and dietary management.[57]

Conclusion

The conclusions that have been reached by this review for urea cycle disorders are similar to those reported by Pollitt and colleagues[1] and Seymour and colleagues.[2] The two existing systematic reviews[1,2] suggest that there are no reliable data available on the expected incidence of urea cycle disorders within the UK. However, the expected incidence of urea cycle disorders, based on estimates from the number of cases clinically diagnosed in the UK (compared with similar disorders of established incidence), is 2.5 cases per 100,000 births. There are uncertainties regarding the effectiveness of treatments for patients with urea cycle defects; however, therapeutic interventions for urea cycle disorders, namely ornithine transcarbamylase deficiency and citrullinaemia, included long-term dietary protein restriction with oral administration of sodium phenylbutyrate. This treatment regimen improved the clinical management of patients with ornithine transcarbamylase deficiency and improved survival rates for patients with citrullinaemia. However, mental retardation, growth retardation, risk of hyperammonaemic episodes and the need for lifetime adherence to strict medication and dietary management accompanied survival for patients with citrullinaemia. No additional evidence was available for the effectiveness of treatments for other urea cycle disorders, such as argininosuccinic aciduria, arginase deficiency and hyperornithinaemia. Evidence regarding the sensitivity and specificity of neonatal screening for urea cycle disorders using tandem MS is limited.

The authors of the Pollitt report[1] reported that the neonatal and later onset forms of urea cycle disorders are life threatening and unless hyperammonaemia can be prevented, mental disability can result. Methods that are available for the rapid removal of ammonia include haemodialysis and haemofiltration. Other methods include the restriction of dietary protein intake with continued oral administration of sodium benzoate, phenylacetate or phenylbutyrate (particularly for citrullinaemia and argininosuccinic aciduria, supplementary arginine). In general, this is the mainstay of long-term treatment for the milder or late-onset forms of urea cycle defects. Infants with the neonatal form of the disease need to be identified and treated within the first week of life; therefore, neonatal screening in the UK between 6 and 10 days of age may be too late. However, patients with the chronic or later onset forms would benefit from the identification and prospective treatment to prevent hyperammonaemia. Data on the sensitivity and specificity of neonatal screening for urea cycle disorders using tandem MS were limited; however, these authors concluded that if methods of sufficient specificity were available, neonatal screening as an add-on test would be justified for patients with chronic or later onset forms of urea cycle defects.

Seymour and colleagues[2] reported that urea cycle defects are associated with significant morbidity and mortality. Therapeutic interventions for urea cycle disorders include the provision of alternative pathways of nitrogen disposal, protein restriction and haemodialysis; however, there were uncertainties regarding the effectiveness of treatments and the period of onset during which intervention improves outcome. For example, the prognosis for urea cycle defects remains generally poor and uncertain even with the best available evidence; however, the prognosis appeared to be excellent for late-onset or partial argininosuccinate lyase deficiency. Neonatal screening in the UK, which is generally performed between 6 and 10 days of age, was unlikely to be of benefit to infants with neonatal onset of urea cycle defects, who will sustain hyperammonaemic brain damage before screening is performed, and the relative proportion of neonatal-onset to late-onset disease was unknown for any of the urea cycle defects. Based on the available evidence, these authors concluded that neonatal screening for urea cycle defects was not justified.

Methylmalonic, propionic and isovaleric acidaemias

Background

Methylmalonic, propionic and isovaleric acidaemias are disorders of organic acid metabolism. Methylmalonic and propionic acidaemias result from inherited defects in the catabolism of propionyl-CoA (a product of the catabolism of certain amino acids and odd-carbon fatty acids). Isovaleric acidaemia is caused by deficient activity of the enzyme isovaleryl-CoA dehydrogenase.

Propionic acidaemia is related to a deficiency of propionyl-CoA carboxylase, resulting in an accumulation of propionic acid and the production of a broad range of abnormal metabolites. Secondary inhibition of other enzyme systems can lead to ketosis, hyperammonaemia

and inhibition of glycine catabolism. Patients with this condition present in early infancy with severe ketoacidosis, hypotonia, arreflexia, vomiting, breathing irregularities, lethargy and coma. There are acute and more chronic forms of the disorder. The acute form may present in the newborn period soon after initial protein feeding. The more chronic form of the disorder may present with recurrent episodes of ketosis, hyperammonaemia and hyperglycinaemia, often precipitated by infection.

There are many types of methylmalonic acidaemia, but most reported cases are caused by a defect in the methylmalonyl-CoA mutase apoenzyme or in the processing of hydroxocobalamin to deoxyadenosyl cobalamin. Again, the inhibition of other enzyme systems can lead to secondary effects such as ketosis, hyperammonaemia and the inhibition of glycine metabolism. Patients with this disorder usually present within the first month of life; they are acutely ill with failure to thrive, hypotonia and ketoacidosis. There may be associated vomiting, dehydration and lethargy, leading to coma and death if left untreated.

Isovaleric acidaemia is caused by deficient activity of the enzyme isovaleryl-CoA dehydrogenase. This leads to a build-up of isovaleryl-CoA in the cells, with the accumulation of the metabolite isovalerylglycine in the body fluids. This produces the characteristic unpleasant odour of patients with this condition. The clinical presentation of this disorder involves episodes of acidosis, vomiting, ataxia and tremors that can progress to coma. Many reported cases are characterised by acute presentation in the first few weeks of life, with severe ketosis and neurological symptoms (convulsions and coma). Further deterioration, with the onset of overwhelming sepsis and leucopenia, can lead to death in approximately half of all such cases.

In addition, two other disorders of organic acid metabolism (3-hydroxy-3-methylglutaryl-CoA lyase deficiency and 3-methylcrotonyl-CoA carboxylase deficiency), which are fairly rare, can be detected using tandem MS. These disorders are reviewed in the next section.

Quantity and quality of research available

Quantity of research available

The systematic searches identified 99 additional publications on the incidence, prevalence and effectiveness of treatment outcomes for methylmalonic, propionic and isovaleric acidaemias. Full copies of 37 papers were retrieved for further evaluation.

Of these 37 papers, two studies were judged to be relevant to the update review. Both publications[58,59] reported information on the effectiveness of treatment outcomes for methylmalonic[58] or propionic[59] acidaemia. No studies were identified that provided data on the incidence of methylmalonic, propionic and isovaleric acidaemias in the UK.

Thirty-five studies were excluded from the update review (details of references are provided in Appendix 24), for the following reasons:

- patients: not isolated methylmalonic defect, i.e. patients with combined methylmalonic acidaemia and homocystinuria ($n = 4$)
- publication type: review ($n = 1$)
- study design: screening ($n = 13$)
- outcome measures: no incidence or prevalence data ($n = 1$), mutation, genotype, etc., analysis ($n = 3$), no treatment outcomes or treatment outcomes in fewer than two patients ($n = 8$), incidence or prevalence data from non-UK countries ($n = 4$)
- other reasons: included in existing systematic reviews ($n = 1$).

Quality of included studies

Incidence of methylmalonic, propionic and isovaleric acidaemias

No additional evidence was available on the incidence of methylmalonic, propionic and isovaleric acidaemias in the UK.

Effectiveness of treatments for methylmalonic, propionic and isovaleric acidaemias (Appendices 25 and 26)

Two studies provided information on the effectiveness of dietary interventions for patients with methylmalonic acidaemia or propionic acidaemia.

In one study[58] the authors evaluated the long-term outcomes of 35 patients with methylmalonic acidaemia using a cross-sectional study design. Patients in this study were seen and treated at the Great Ormond Street Hospital for Children, London, UK, between 1970 and 1996. All patients were diagnosed with methylmalonic acidaemia on the basis of increased urinary methylmalonate and methylcitrate, raised methylmalonate in blood with normal plasma vitamin B_{12} levels and no detectable plasma homocystine. In most patients

diagnosis was confirmed by enzyme studies on cultured skin fibroblasts. After diagnosis all patients were treated with a low-protein diet and subsequently given a trial of intramuscular injections of cyanocobalamin or hydroxycobalamin (1 mg daily for 5 days). The response to this was assessed by urinary methylmalonate measurements and in the responsive patients cobalamin injections were continued. Consequently, patients were divided into cobalamin-responsive (n = 6) and non-responsive (n = 29) and early-onset (presented in the first month of life) and late-onset (presented after the first month) groups.

It is not explicitly clear how many patients were lost to follow-up; however, 14 patients in the early-onset cobalamin-non-responsive group died. The authors did not report the duration of the study, inclusion/exclusion criteria, mean age of patients (however, all patients were under 16 years of age), compliance to treatment or gender characteristics.

Van der Meer and colleagues[59] conducted a retrospective analysis of clinical data to evaluate the clinical outcome and long-term treatment of 17 patients with propionic acidaemia. Patients in this study were diagnosed with propionic acidaemia in the early 1970s, with early-onset (n = 12) or late-onset type (n = 5). In most patients, diagnosis was confirmed with an enzymic assay of propionyl CoA carboxylase. The biotin response was also assessed in all patients, but none responded either clinically or biochemically. Twelve patients presented within 3 weeks after birth (mean 9.3 days, range 3–19 days; early-onset group), whereas five patients were diagnosed later in life (mean 16.3 months, range 3.5 months to 3 years of age; late-onset group). At the time of the study, the median age of the living early-onset patients was 5.2 years (range 1–9.3 years), whereas the late onset patients were 4, 7 and 23 years of age.

The basic principles for the dietary treatment of patients with propionic acidaemia involved dietary management with low protein intake (natural) with or without supplemental protein mixtures. In the 1970s dietary treatments for propionic acidaemia were based on restriction of natural protein. After 1980, the treatment principles changed. Home tube-feeding became a routine daily treatment, with the addition of carnitine (100 mg/kg per day) and later metronidazole (20 mg/kg per day). All parents were informed and trained in how to manage the patient under circumstances such as fever, vomiting and food refusal, and in how to place a nasogastric tube in case of urgent or semiurgent treatment at home. The changes in regimen were noted and kept under strict surveillance.

It is not explicitly clear how many subjects were lost to follow-up; however, seven patients died (five in the early-onset group and two in the late-onset group) during the study period. The authors did not report the duration of the study, inclusion/exclusion criteria, compliance with treatment or gender characteristics.

The strength and quality of the evidence for the effectiveness of dietary and/or pharmacological interventions for methylmalonic acidaemia and propionic acidaemia were grade III.[58,59]

Summary of results

Incidence of methylmalonic, propionic and isovaleric acidaemias

No additional evidence was available on the incidence of methylmalonic, propionic and isovaleric acidaemias in the UK. Owing to the lack of additional evidence, incidence data from the existing systematic reviews[1,2] have been reported.

Pollitt and colleagues[1] reported that the collective incidence of methylmalonic, propionic and isovaleric acidaemias was three cases per 100,000 births in the UK. Seymour and colleagues[2] reported that the individual incidence for disorders of organic acid metabolism may be low; however, collectively (methylmalonic aciduria, propionic acidaemia, isovaleric acidaemia, 3-hydroxy-3-methylglutaric aciduria and glutaric aciduria type I) the incidence may be estimated at around 1:15,000 live births.

Effectiveness of treatments for methylmalonic, propionic and isovaleric acidaemias

A cross-sectional study[58] that assessed the long-term outcomes of 35 patients with methylmalonic acidaemia found significant differences between cobalamin responsive and non-responsive groups in the severity, survival and incidence of neurological sequelae.

All six cobalamin-responsive patients had mild disease and its severity was not dependent on age at presentation, their neurological complications were less severe and they were all alive. In contrast, the cobalamin-non-responsive group consisted of 19 early-onset and nine late-onset patients. The early-onset group had more severe disease presentation with episodes of severe acidosis and hyperammonaemia (up to 1800 μM), and six collapsed, requiring intensive care and

ventilatory support. Fourteen patients in the early-onset cobalamin-non-responsive group had died (median survival 6.4 years, 95% CI 3.6 to 9.1 years), whereas all the patients in the late-onset group were alive. Six patients in the early-onset cobalamin-non-responsive group died in the first year of life and eight died between 15 months and 7 years of age. Most late-onset patients presented in the first year with either an episode of metabolic decompensation and subsequent development delay, or feeding difficulties with failure to thrive. Anorexia and feeding difficulties were also seen in 25 of the 27 of cobalamin-non-responsive patients.

There were no significant differences between the groups with respect to protein intake, which varied between 1.0 and 1.8 g/kg per day. One-third of cobalamin-non-responsive patients had poor growth with height and weight below the third centile. No significant differences were seen between the early- and late-onset groups in abnormal neurological signs (seven of 18 and six of nine, respectively, $p = 0.13$), although early-onset patients had significantly reduced full-scale IQ (median difference 26, $p = 0.03$) and poor neurological and cognitive outcome. In both groups abnormal neurological signs continued to increase with age. In all patients, new neurological symptoms and signs developed following episodes of acute metabolic decompensation.

In the long-term treatment of 17 patients with propionic acidaemia,[59] five (42%) of the early-onset and two (40%) of the late-onset patients died. The deceased early-onset patients had a median survival time of 0.4 years, whereas the late-onset patients died at the ages of 2.8 and 4 years. At the time of study, the median age of the living early-onset patients was 5.2 years (range 1–9.3 years), while the late-onset patients were 4, 7 and 23 years of age.

All patients were treated with natural protein restriction with addition of carnitine (100 mg/kg per day) and later with metronidazole (20 mg/kg per day). The early-onset patients were almost all treated with daily home tube-feeding. The natural protein intake per day remained fairly stable during the first 3–4 years of life in the early-onset patients. After the sixth year of life, total protein intake remained fairly constant and seldom reached values greater than 13 g/day. The supplemental protein intake was higher in early-onset patients and showed a steady but strong increase and levelled off from the age of 6 years. In contrast, the natural protein intake of late-onset patients was higher and rose more rapidly to an almost normal protein intake after 3–4 years of life. The differences between the mean natural protein intake of early-onset and late-onset patients were significant in all age groups ($p < 0.0001$, *t*-test).

Late-onset patients suffered more frequently from minor to intermediate neurometor, mental and psychological disabilities than the early-onset patients; this may be due to the delay in diagnosis. The authors also observed less frequent metabolic decompensations and hospitalisations with the introduction of nasogastric tube feeding. Overall, many patients showed a failure to thrive, particularly for height. This was attributed to beginning strong protein restriction during the first years of life.

Discussion

Despite the lack of additional evidence on the incidence of methylmalonic, propionic and isovaleric acidaemias in the UK, information reported in the Pollitt review[1] suggests a collective incidence of three cases per 100,000 births in the UK. However, these findings are primarily based on extrapolated data from non-UK populations and from the number of cases clinically diagnosed in the UK (compared with similar disorders of established incidence). In contrast, Seymour and colleagues[2] reported that the collective incidence for the disorders of organic acid metabolism was estimated to be around 1:15,000 live births. This figure also included other defects of organic acid metabolism (methylmalonic aciduria, propionic acidaemia, isovaleric acidaemia, 3-hydroxy-3-methylglutaric aciduria and glutaric aciduria type I), which have been covered elsewhere in the review. Overall, there appears to be a lack of individual data for the incidence of methylmalonic, propionic and isovaleric acidaemias in the UK.

The evidence for the effectiveness of dietary and or pharmacological interventions in patients with methylmalonic or propionic acidaemia was limited to a few studies. Despite the poor study quality, the results showed that the overall outcome of patients with methylmalonic acidaemia,[58] particularly the early-onset group, was disappointing and unsatisfactory. In this group, patients had more severe disease at presentation and 14 patients died during the study period. All cobalamin-non-responsive patients were at risk of developing a progressive neurological disease, particularly the early-onset group. In general, cobalamin-responsive patients had a better long-term outcome than non-cobalamin, responsive

patients. The therapeutic outcomes for patients with propionic acidaemia[59] appeared to be satisfactory in terms of survival, neurological and mental development; however, in this study five (42%) of the early-onset and two (40%) of the late-onset patients died during the study period. The mean natural protein intake of early-onset patients was significantly lower than that of late-onset patients. Supplemental protein intake was also higher in early-onset patients. The general neurological outcomes were generally better for early-onset patients and most patients encountered growth retardation. Overall, the burden on the patient, the parents and the medical profession is amplified for the adequate management of this defect.

Conclusion

The conclusions that have been reached by this review for methylmalonic or propionic acidaemia are similar to those reported by Pollitt and colleagues[1] and Seymour and colleagues.[2] The two existing systematic reviews[1,2] suggest that there is a lack of individual data for the incidence of methylmalonic, propionic and isovaleric acidaemia in the UK. However, the collective incidence was estimated to be three cases per 100,000 births in the UK. Methylmalonic and propionic acidaemia were associated with increased morbidity and mortality and there were uncertainties regarding the effectiveness of treatments for patients with methylmalonic and propionic acidaemia, especially with the neonatal-onset form. The mainstay of long-term therapy for these disorders appeared to be dietary protein restriction with or without cobalamin therapy or supplemental protein mixtures. No additional evidence was available for the effectiveness of treatments for isovaleric acidaemia. Evidence regarding the sensitivity and specificity of neonatal screening for methylmalonic, propionic and isovaleric acidaemia using tandem MS is limited.

The authors of the Pollitt review[1] reported that neonatal screening was unlikely to be of benefit to patients with the neonatal-onset forms of methylmalonic, propionic or isovaleric acidaemia; however, the later onset variants were more amenable to treatment. These disorders were associated with significant mortality and morbidity. Long-term management involved dietary protein restriction, with or without a combination of artificial amino acid supplements, particularly for the severe forms of methylmalonic and propionic acidaemia. Other therapeutic regimens included cobalamin therapy (not all methylmalonic patients are responsive), substrate removal (glycine is used in isovaleric acidaemia, carnitine is used for organic acidaemias, but its effectiveness on outcome is controversial) and liver transplantation (only a few successful treatments have been reported). Overall, the effectiveness of treatment was low to high. Data on the sensitivity and specificity of neonatal screening for methylmalonic, propionic or isovaleric acidaemia using tandem MS were limited; however, these authors reported that propionylcarnitine and isovalerylcarnitine could be readily quantified with tandem MS, giving diagnoses of propionic and isovaleric acidaemia. The diagnosis of methylmalonic acidaemia relies on detecting the secondary accumulation of propionylcarnitine, because methylmalonylcarnitine is not readily detected using tandem MS. Furthermore, mild cases of methylmalonic acidaemia are unlikely to be detected by neonatal screening.

Seymour and colleagues[2] reported that methylmalonic, propionic and isovaleric acidaemias were associated with significant morbidity and mortality. Therapeutic interventions for patients with disorders of organic acid metabolism include long-term management involving substrate restriction though dietary manipulation, the use of cofactors at the pharmacological level, removal of substrate by conjugation or inhibition of endogenous synthesis, and the use of L-carnitine. These authors found that effective treatments were available for isovaleric acidaemia, further research was needed for methylmalonic acidaemia and there were uncertainties regarding the effectiveness of treatments for propionic acidaemia. Furthermore, findings were ambiguous regarding the period before onset during which intervention improved outcome. The use of tandem MS for neonatal screening for acylcarnitines in blood spots appeared to be a viable procedure for the identification of patients with methylmalonic, propionic or isovaleric acidaemia; however, taken individually there was no justification for screening for them at the present time. Nevertheless, collectively, these disorders can be detected at no additional screening cost, there would be clinical benefits for some of these patients and early identification would assist genetic counselling and prevent further cases in the same families.

Other defects of branched-chain acyl-CoA metabolism

Background

This section deals briefly with two disorders, which are fairly rare but can be detected in newborns by

acylcarnitine analysis in dried blood spots using tandem MS.[6] Both 3-hydroxy-3-methylglutaryl-CoA lyase deficiency and 3-methylcrotonyl-CoA carboxylase deficiency affect the pathway of leucine metabolism. 3-Hydroxy-3-methylglutaryl-CoA lyase deficiency presents acutely in the neonate or more frequently from 3–11 months. In both cases the disorder is characterised by lethargy, hypotonia, vomiting, metabolic acidosis, hypoglycaemia and hepatomegaly with hyperammonaemia. In addition to affecting the pathway of leucine metabolism, the enzyme is involved in ketone body production and utilisation and is therefore important in the physiological response to prolonged fasting.[1] 3-Methylcrotonyl-CoA carboxylase deficiency is a less severe condition usually presenting in the second or third year of life. It is characterised by a Reye-like illness following infection, with persistent hypotonia as a presenting symptom.

Quantity and quality of research available

Quantity of research available

The yield of literature from the systematic searches generated 51 additional publications on the incidence, prevalence and effectiveness of treatment outcomes for 3-hydroxy-3-methylglutaryl-CoA lyase deficiency and 3-methylcrotonyl-CoA carboxylase deficiency. Full copies of nine papers were retrieved for further evaluation.

Of these nine papers, none was judged to be relevant to the update review. These nine studies were excluded from the update review (details of references are provided in Appendix 27), for the following reasons:

- study design: screening (n = 2)
- outcome measures: mutation, genotype, etc., analysis (n = 5), treatment outcomes in fewer than two patients (n = 1), incidence or prevalence data from non-UK country (n = 1).

Quality of included studies

Incidence of other defects of branched-chain acyl-CoA metabolism

No additional evidence was available on the incidence of 3-hydroxy-3-methylglutaryl-CoA lyase deficiency and 3-methylcrotonyl-CoA carboxylase deficiency in the UK.

Effectiveness of treatments for other defects of branched-chain acyl-CoA metabolism

No additional evidence was available on the effectiveness of treatments for 3-hydroxy-3-methylglutaryl-CoA lyase deficiency and 3-methylcrotonyl-CoA carboxylase deficiency.

Summary of results

Incidence of other defects of branched-chain acyl-CoA metabolism

No additional evidence was available on the incidence of 3-hydroxy-3-methylglutaryl-CoA lyase deficiency and 3-methylcrotonyl-CoA carboxylase deficiency in the UK. Pollitt and colleagues[1] reported that there were no data on the prevalence of these disorders in the UK. Collectively, these disorders were relatively rare and the expected incidence, based on estimates from the number of cases clinically diagnosed in the UK (compared with similar disorders of established incidence), was one case per 100,000 births. In contrast, Seymour and colleagues[2] reported that the UK incidence of 3-hydroxy-3-methylglutaryl-CoA lyase deficiency was unknown.

Effectiveness of treatments for other defects of branched-chain acyl-CoA metabolism

No additional evidence was available on the effectiveness of treatments for 3-hydroxy-3-methylglutaryl-CoA lyase deficiency and 3-methylcrotonyl-CoA carboxylase deficiency. Despite the lack of additional evidence on the effectiveness of treatments, data reported in the Pollitt review[1] suggest the availability of various treatments. For these disorders, long-term therapy consists of dietary protein restriction, avoidance of prolonged fasting and sometimes oral supplementation of L-carnitine. A fat-restricted diet is also recommended for patients with 3-hydroxy-3-methylglutaryl-CoA lyase deficiency. Acute attacks are treated by intravenous glucose and correction of acidosis.

Discussion

No additional evidence was available on the UK incidence and effectiveness of treatments for 3-hydroxy-3-methylglutaryl-CoA lyase deficiency and 3-methylcrotonyl-CoA carboxylase deficiency. Information for these disorders was derived solely from the two existing systematic reviews.[1,2]

Conclusion

Despite the lack of additional evidence on the incidence of 3-hydroxy-3-methylglutaryl-CoA lyase deficiency and 3-methylcrotonyl-CoA carboxylase deficiency in the UK, data reported in the Pollitt and Seymour reviews suggest that the collective incidence of these disorders is relatively low.[1,2] Pollitt and colleagues[1] estimated the collective incidence as one case per 100,000 births. Evidence is very limited regarding the sensitivity and specificity of neonatal screening for 3-hydroxy-3-methylglutaryl-CoA lyase deficiency and 3-methylcrotonyl-CoA carboxylase deficiency using tandem MS.

Pollitt and colleagues[1] concluded that the presence of asymptomatic cases raises similar problems to patients with MCAD deficiency. The potentially serious effect of these conditions makes them suitable for neonatal screening; however, there are insufficient data for these disorders to predict the long-term outlook in any detail.

Seymour and colleagues[2] found that 3-hydroxy-3-methylglutaryl-CoA lyase deficiency was associated with significant morbidity and mortality and that effective treatments were available for this disorder. However, there were uncertainties regarding the period before onset during which the intervention improved outcome. These authors concluded that other disorders of organic acid metabolism should not be considered for newborn screening at the present time.

MCAD

Background

MCAD deficiency is one of the most common fatty acid oxidation disorders. Patients may present acutely with hyperammonaemia, hypoglycaemia, encephalopathy and hepatomegaly, or mildly with an isolated hypoglycaemic episode. The age of first presentation varies widely from early infancy to later childhood, or may even occur in an adult with hypoglycaemic episodes or muscle weakness. This range of severity and first presentation is related to the deficiency of medium-chain acyl-CoA dehydrogenase causing only partial blockage of fatty acid oxidation. This occurs because there is some degree of overlap between some of the chain-length specific enzymes. Therefore, the disorder may only become manifest when demands on fatty acid oxidation are particularly high, for example, during a period of fasting that is often related to an infectious illness where there has been reduced food intake.

Although the severity of metabolic crises related to this disorder is very variable, it is more usual for attacks to be progressive and severe rather than isolated and mild. Untreated metabolic crises can lead to seizures, apnoea and cardiac arrest. In some cases death has occurred overnight and has been ascribed to sudden infant death syndrome (SIDS). In general, studies have established that there is significant mortality (20–25%) and morbidity from the initial metabolic crisis in previously undiagnosed cases of MCAD deficiency.[60–63] However, some MCAD-deficient individuals may remain asymptomatic throughout life.[64]

Quantity and quality of research available

Quantity of research available

The systematic searches identified 108 additional publications on the incidence, prevalence and effectiveness of treatment outcomes for MCAD deficiency. Full copies of 50 papers were retrieved for further evaluation.

Of these 50 papers, three studies were judged to be relevant to the update review. All three publications[21,30,65] provided data on the incidence of MCAD deficiency in the UK. No studies were identified that reported information on the effectiveness of treatment outcomes for MCAD deficiency.

Forty-seven studies were excluded from the update review (details of references are provided in Appendix 28), for the following reasons:

- publication type: letters, comments or statements ($n = 3$), review ($n = 4$)
- study design: screening ($n = 12$)
- outcome measures: no data on incidence or prevalence ($n = 1$), no treatment outcomes or treatment outcomes in fewer than two patients ($n = 4$), incidence or prevalence data from non-UK countries ($n = 19$)
- other reasons: same results reported elsewhere ($n = 1$), included in existing systematic reviews ($n = 3$).

Quality of included studies

Incidence of MCAD deficiency (Appendix 29)

Three studies provided data on the incidence of MCAD deficiency in the UK.[21,30,65] Two studies used a retrospective cohort design and evaluated all newborn infants in the West Midlands NHS region during the period 1981–91[30] and in the Northern NHS region during the period 1991–3.[21] In contrast, Pollitt and Leonard[65] used a prospective surveillance study through the British Paediatric Surveillance Unit to investigate the diagnosis and outcome of MCAD deficiency in the UK between 1994 and 1996.

Retrospective data for the inherited metabolic disorders in the Hutchesson[30] study were derived from the West Midlands Neonatal Screening Programme, the regional register for patients with inborn errors of metabolism and population frequencies from the national census. The number of patients with confirmed inherited metabolic disorders born in the region was derived from laboratory records and their ethnicity from personal knowledge through involvement with

clinical management (further details of quality assessment were discussed in Incidence of PKU, towards the beginning of this chapter).

Pourfarzam and colleagues[21] analysed the concentrations of acylcarnitines in stored neonatal blood spots (up to 5 years) and reviewed patients with high octanoylcarnitine concentrations at the age 7–9 years. In this study, many cases of MCAD deficiency were not diagnosed on clinical grounds.

In the Pollitt and Leonard[65] prospective study, the British Paediatric Surveillance Unit sent out orange cards on a monthly basis to all consultant paediatricians in the UK and the Republic of Ireland, accompanied by short case definitions of the disorders under surveillance. All notifications were followed up by a brief questionnaire requesting patient details, presentation, family history and diagnostic criteria. This information was supplemented by data from UK laboratories that were likely to have diagnosed and confirmed MCAD deficiency.

The authors of these UK studies failed to report any details on the age of sampling,[21,65] gender[21,30,65] or ethnic[21,65] characteristics.

The strength and quality of the evidence for the expected incidence of MCAD deficiency within the UK were grade I[65] and grade Ia.[21,30]

Effectiveness of treatments for MCAD deficiency
No additional evidence was available on the effectiveness of treatments for MCAD deficiency.

Summary of results

Incidence of MCAD deficiency in the UK

Study details and results are summarised in *Table 9*.

The prevalence of diagnosed MCAD deficiency among births in the West Midlands NHS region during 1981–91 was 1.3 per 100,000 births.[30] In the Northern NHS region during 1991–3, the prevalence of diagnosed cases was somewhat higher, 3.0 per 100,000 births,[21] which could reflect either geographical differences in prevalence or improved clinical recognition. The true birth prevalence or incidence of MCAD deficiency in the Northern region was 8.0 per 100,000 births, since the majority of cases typically go undiagnosed.

Between 1994 and 1996, the prospective surveillance study found 62 affected individuals with MCAD deficiency from 54 families in the UK.[65] Of these 62 diagnosed cases, 57 were from England (prevalence of diagnosed MCAD deficiency, 4.5 cases per 100,000 births) and five from Scotland (prevalence of diagnosed MCAD deficiency, 4.0 cases per 100,000 births). No reports were received from Wales, Northern Ireland or the Republic of Ireland. The authors[65] reported that there were no firm data on the proportion of cases that remained undiagnosed or asymptomatic; nevertheless, based on these data, the minimum true birth prevalence or incidence of MCAD deficiency in the UK was 4.0 per 100,000 births.

Hutchesson and colleagues[30] found that the prevalence of diagnosed MCAD deficiency was appreciably higher in north-west Europeans (1:67,259, 95% CI 1:35,430 to 1:147,089) than in the Pakistani group (0, 95% CI < 1:9350). These authors also observed that the gene frequency for MCAD deficiency was significantly lower in the Pakistani population (0, 95% CI < 1:395) than in north-west European children (1:259, 95% CI 1:188 to 1:384, $p < 0.05$), indicating that the diagnosis for MCAD deficiency was rare outside those of north-west European ethnicity.

In the Pourfarzam study,[21] seven of the eight MCAD-deficient patients were homozygous and one was heterozygous for the 985A → G mutation. This mutation accounted for 94% of mutant alleles. These retrospective findings were consistent with those reported by Pollitt and Leonard.[65] In this UK prospective study of clinically detected MCAD-deficient patients, DNA analysis for the common 985A → G mutation in 45 families revealed that 36 affected children were homozygous for the mutation, whereas nine were heterozygous. The 985A → G mutation in this study accounted for 90% of mutant alleles.

Effectiveness of treatment for MCAD deficiency

No additional evidence was available on the effectiveness of treatments for MCAD deficiency (see following 'Discussion' section).

Discussion

The evidence for MCAD deficiency is limited, but suggests a relatively high true birth prevalence/incidence and wide geographical variation in the frequency of MCAD deficiency in the UK. The prevalence figures in the Hutchesson[30] study may be an underestimate because the authors reported that they were unable to identify those who died before collection of the neonatal sample and children with inborn errors may have died without recognition of the underlying diagnosis, particularly

TABLE 9 *Frequency of MCAD deficiency in the UK*

Study	Location	Study design	Period of study	Patient selection	Age at sampling	Confirmation of disease	Population screened	Cases diagnosed	Cumulative frequency	Cases per 100,000	Quality of evidence (grade)
Hutchesson *et al.*[30]	Neonatal Screening Service Programme, West Midlands, UK	Retrospective cohort study	April 1981 to April 1991	All newborns in West Midlands NHS region	NR	Yes, laboratory records	707,720	9[a]	9:707,720	1.3	Ia
Pourfarzam *et al.*[21]	Northern region of the NHS, UK	Retrospective, cohort study	January 1991 to July 1993	All newborns in Northern NHS region	NR	Yes, used urinary markers and DNA analysis for mutations (985A → G)	100,600 (tandem MS-based screening)	8[b]	8:100,600	8.0	Ia
Pollitt and Leonard[65]	British Paediatric Surveillance Unit, UK	Prospective, cohort study	March 1994 to March 1996	All patients diagnosed with MCAD deficiency in the UK	NR	Yes, laboratory and clinical presentation	NR	62[c]	NR	England 4.5 Scotland 4.0	I

[a] Disease frequency data for different ethnic groups [cases per 100,000 (95% CI)]: north-west European, 1:67,259 (1:35,430 to 1:147,089); Pakistani, 0 (< 1:9350).
[b] One patient died before diagnosis, four had life-threatening illnesses, one had mild symptoms and two had no symptoms.
[c] Fifteen patients were asymptomatic, 30 recovered from an attack, six showed signs of neurological impairment or developmental delay, ten died from the initial attack and no information was available on one patient.
NR, not reported.

those with MCAD deficiency, which can present with neonatal death. Moreover, some may have failed as yet to present clinically, particularly those with MCAD deficiency, which is frequently underdiagnosed.[30]

Almost universally, the number of cases diagnosed is substantially lower than the calculated incidence. Asymptomatic individuals and/or low rates of diagnosis of symptomatic cases account for this shortfall, but the relative magnitude of these effects is unknown.[65] In the absence of population-based screening data in the UK, studies of gene frequencies can be used to calculate the expected incidence of MCAD deficiency. At least 90% of clinical cases of MCAD deficiency in the UK are homozygotes for the 985A → G mutation,[21,65,66] but there is substantial geographical variation in the frequency of this mutation. No national-level representative data are available; however, estimates from two English health regions for the common 985A → G mutation (data included in the Pollitt report[1]) found that in the Trent and West Midlands Health Region combined, one in 64 of the newborn population is heterozygous for the 985A → G mutation. The calculated incidence of MCAD deficiency in the population studied (all mutations, assuming that 90% of MCAD mutations are 985A → G) was 7.4 cases per 100,000 births (95% CI, 5.3 to 9.9 per 100,000).[66] Those data, together with retrospective and prospective data presented in this review, suggest that the true expected birth prevalence or incidence of MCAD deficiency in the UK ranges between 4.0 and 9.9 cases per 100,000 births. In some studies detailed follow-up has shown that the proportion of the common MCAD mutation is less than expected and patients are being detected with a mutation that has never been found to present clinically.[23]

Despite the lack of additional evidence on the effectiveness of treatments for MCAD deficiency, various studies of screened cohorts (tandem MS-based newborn screening) have consistently reported no deaths[20,23] in infants found to have MCAD deficiency, with the exception of the first two infants detected by tandem MS in Pennsylvania, USA.[23,67] However, this was before the institution of adequate clinical follow-up[63,67] and since then there have been no further deaths among patients diagnosed with MCAD deficiency in Pennsylvania.[28] In a before and after study that investigated the outcome of MCAD deficiency after established diagnosis, Wilson and colleagues[63] found that from the time of diagnosis and initiation of an effective yet relatively inexpensive treatment (avoidance of fasting and the use of regular high carbohydrate drinks during periods of infection and anorexia) there were no deaths or appreciable additional cognitive impairment.

In a prospective surveillance study, which reported the rate of diagnosed MCAD cases and their associated outcomes in an unscreened population, Pollitt and Leonard[65] found 62 diagnosed cases of MCAD-deficiency. Of these 62 diagnosed cases, 46 MCAD-deficient patients presented with an acute illness; 13 were identified because of family history and three for other reasons. The outcomes for the MCAD-deficient patients at the time of study were death from acute illness (16%), neurological impairment or developmental delay (10%), full recovery from attack (48%) and no symptoms (24%), and no outcome information was available for one patient (2%). A longer follow-up period would be necessary to reveal the milder neurodevelopmental deficits and since follow-up has not been conducted past childhood or past the immediate postillness period, it is unknown whether this number for neurological impairment or developmental delay would increase if follow-up were performed through adulthood. All but two of the clinically affected cases presented with typical symptoms of MCAD deficiency. Although a few were diagnosed in the neonatal period, most patients had their first detected attack after the age of 3 months. The age at first episode ranged from 2 days to 4.39 years, with a median at 1.1 years. Neonatal episodes may have been significantly underreported because there was no methodical review of the neonatal records.[65]

In patients whose disorder was diagnosed by retrospective screening of stored neonatal blood spots, Pourfarzam and colleagues[21] detected eight cases of MCAD deficiency by tandem MS. Of the eight detected cases of MCAD deficiency one patient died of gastroenteritis at age 17 months, before diagnosis; four others had life-threatening illnesses (one had neonatal apnoea related to MCAD deficiency and three had severe or recurrent episodes of encephalopathy), one had mild symptoms of MCAD deficiency and two had no symptoms (25%). Most patients in this study with MCAD deficiency were not diagnosed on clinical grounds but developed symptoms in early childhood. Only one patient was severely symptomatic in the neonatal period.

Conclusion

The conclusions that have been reached by this review for MCAD deficiency are very similar to those reported by Pollitt and colleagues[1] and Seymour and colleagues.[2] Despite the limitations

of the UK studies, the true expected birth prevalence or incidence of MCAD deficiency in the UK ranges between 4.0 and 9.9 cases per 100,000 births and is more common among north-west Europeans than other ethnic groups. Discrepancies between the incidence predicted from the gene frequency data and the number of patients diagnosed arise because MCAD deficiency can present with neonatal death, and an unknown number of patients may remain asymptomatic or have mild symptoms and may never experience any ill effects and difficulties arising in clinical diagnoses. The disorder has been shown to be associated with high morbidity and mortality; however, after (early) diagnosis, current management makes death rare and improves outcome.[63] Neonatal screening data from tandem MS-based studies have shown that this method is robust, highly sensitive (100%) and specific (100%) for MCAD deficiency.

The authors of the Pollitt report[1] found that the expected incidence of MCAD deficiency in the UK is 5–10 cases per 100,000 births, treatment is relatively simple with dietary management directed primarily towards preventing fasting stress and the overall sensitivity of tandem MS screening is >95% for MCAD deficiency. These authors concluded that MCAD deficiency fulfils all of the classical requirements for a screening programme; however, the main concern is that a proportion of biochemically affected individuals will be asymptomatic and will never experience any ill effects, and there is no way of predicting who these will be. Therefore, all neonates with MCAD deficiency detected by newborn screening must be treated as being at risk and the need to monitor intercurrent illnesses for the first few years of life will cause parental anxiety. There is also the possibility that "a good counselling and support service with experienced clinicians can reduce this" (Leonard J: personal communication; 2002).

Seymour and colleagues[2] found that MCAD deficiency showed a high frequency in the general population (estimated incidence 1:8000 to 1:15,000), especially in Caucasians of north European descent, and is associated with increased morbidity and mortality. MCAD deficiency is simply and easily treated with simple dietary manipulation and L-carnitine supplementation, thus preventing possible early death and neurological disability. These authors concluded that screening for MCAD deficiency should be seriously considered for inclusion in neonatal screening programmes, but such screening is dependent on the introduction of tandem MS. Individuals with MCAD deficiency who may (always) remain asymptomatic and the possibility of (even minor) clinically unwarranted treatment present important considerations and challenges for neonatal screening programmes for MCAD deficiency.

Defects of long-chain fatty acid catabolism

Background

Several disorders caused by defect in the transport and degradation of long-chain fatty acids have been identified in relation to deficiency of individual enzymes: carnitine palmitoyltransferase type I, carnitine-acylcarnitine translocase, carnitine palmitoyltransferase type II, very long-chain acyl-CoA dehydrogenase, long-chain enoyl-CoA hydratase, long-chain 3-hydroxyacyl-CoA dehydrogenase and long-chain thiolase.

The clinical presentation and severity of all these conditions vary, but they share a general pattern. Apart from carnitine palmitoyltransferase type II deficiency, which can cause congenital defects and prove rapidly fatal, these disorders produce a range of presenting symptoms similar to those of MCAD deficiency but with greater severity, earlier onset and additional features. In the neonate there is a tendency towards cardiac symptoms and cardiac arrest, with hypoglycaemic attacks a less consistent feature. Whereas chronic presentations are rare in MCAD deficiency, they are more likely to be a feature of long-chain defects, producing failure to thrive, recurrent vomiting, diarrhoea, hypotonia or cardiomyopathy.

Quantity and quality of research available

Quantity of research available

The systematic searches identified 245 additional publications on the incidence, prevalence and effectiveness of treatment outcomes for the defects of long-chain fatty acid catabolism. Full copies of 30 papers were retrieved for further evaluation.

Of these 30 papers, none was judged to be relevant to the update review. These studies were excluded from the update review (details of references are provided in Appendix 30), for the following reasons:

- publication type: review (n = 2)
- study design: screening (n = 11), case report and survey of physicians (n = 1)

- outcome measures: mutation, genotype, etc., analysis (n = 7), no treatment outcomes or treatment outcomes in fewer than two patients (n = 5), incidence or prevalence data from non-UK countries (n = 3)
- other reasons: same results reported elsewhere (n = 1).

Quality of included studies

Incidence for defects of long-chain fatty acid catabolism

No additional incidence data were available for defects of long-chain fatty acid catabolism in the UK.

Effectiveness of treatments for defects of long-chain fatty acid catabolism

No additional evidence was available on the effectiveness of treatments for defects of long-chain fatty acid catabolism.

Summary of results

Incidence for defects of long-chain fatty acid catabolism

No additional incidence data were available for defects of long-chain fatty acid catabolism within the UK. Owing to the lack of additional evidence, incidence data from the existing systematic reviews[1,2] have been reported.

Pollitt and colleagues[1] and Seymour and colleagues[2] reported that the true incidence for these disorders was unknown and there was no secure data available on the prevalence. This is because these disorders have been identified relatively recently, clinical presentations are varied and non-specific, laboratory diagnosis is often difficult even in specialised centres, and there is limited information on the precise biochemical basis, natural history and prognosis of the disorders and the availability of effective treatments. Based on estimates from the number of cases clinically diagnosed in the UK (compared with similar disorders of established incidence), Pollitt and colleagues[1] estimated the expected incidence as three cases per 100,000 births.

Effectiveness of treatments for defects of long-chain fatty acid catabolism

No additional evidence was available on the effectiveness of treatments for defects of long-chain fatty acid catabolism. Pollitt and colleagues[1] reported that normally a carbohydrate-rich diet is prescribed, with restriction of long-chain fats and their replacement by medium-chain triglycerides for patients with defects of long-chain fatty acid catabolism. Similar findings were reported by Seymour and colleagues.[2] This strict dietary regimen has been shown to be highly effective in a number of patients; however, based on limited experience it appears that carnitine-acylcarnitine translocase deficiency is essentially untreatable, at least in its more severe form, and there was no long-term experience of outcome for disorders of long-chain fatty acid oxidation.[1]

Discussion

No additional evidence was available on the UK incidence and effectiveness of treatments for defects of long-chain fatty acid catabolism. Information for these disorders was derived solely from the two existing systematic reviews.[1,2]

Conclusion

Despite the lack of additional evidence on the incidence for defects of long-chain fatty acid catabolism within the UK, data reported in the Pollitt review[1] suggest that the collective incidence for defects of long-chain fatty acid catabolism was three cases per 100,000 births. Evidence is very limited regarding the sensitivity and specificity of neonatal screening for defects of long-chain fatty acid catabolism using tandem MS.

Pollitt and colleagues[1] concluded that collectively, these diseases were relatively common and were life threatening or produced severe chronic disability. Therapeutic interventions were reasonably effective and case finding by neonatal screening would seem to be justified even though sensitivity has yet to be established; however, for screening to be practicable, specificity would need to be high.

Seymour and colleagues[2] concluded that specific directed newborn screening for defects of long-chain fatty acid catabolism cannot be advocated owing to limited knowledge on the incidence, natural history, treatment and outcome of the disorders. Tandem MS-based screening is unlikely to detect all affected individuals owing to the difficulties in the comprehensive detection (long-chain acylcarnitines may not accumulate in the early neonatal period in all of these conditions and therefore low-level accumulation may be difficult to detect) and interpretation of the results.

Glutaryl-CoA dehydrogenase deficiency

Background

Glutaryl-CoA dehydrogenase deficiency, often referred to as glutaric acidaemia type I or glutaric aciduria type I (GAI), produces inhibition of the

catabolic pathways of lysine and tryptophan. This leads to an accumulation of glutaric acids in the tissues and body fluids. Typically, the disorder presents acutely in a child who has appeared well or shown only mild neurological symptoms. A frequent presenting sign of the disorder is excessive increase in the circumference of the head with associated frontotemporal atrophy. Acute attacks tend to be triggered by infections and the susceptibility to attacks tends to decrease with maturity. Irreversible loss of motor function involving severe dystonia and dyskinesis may result from an acute attack.

Diagnosis may be difficult in symptomatic cases and may be confused with cerebral palsy or poliomyelitis. Untreated patients experience severe progressive neurological deterioration with recurrent attacks leading eventually to death. Evidence suggests that individuals with symptomatic presentation of GAI have high rates of mortality and significant levels of neurological disability, although there remains some uncertainty about the value of particular elements of current therapeutic regimens.[68–70]

Quantity and quality of research available

Quantity of research available

The systematic searches identified 97 additional publications on the incidence, prevalence and effectiveness of treatment outcomes for GAI. Full copies of 21 papers were retrieved for further evaluation.

Of these 21 papers, two studies were judged to be relevant to the update review. Both studies[69,71] provided information on the effectiveness of treatment outcomes for patients with GAI. No studies were identified that reported data on the incidence of GAI in the UK

Nineteen studies were excluded from the update review (details of references are provided in Appendix 31), for the following reasons:

- publication type: review (n = 4)
- study design: screening (n = 4)
- outcome measures: mutation, genotype, etc., analysis (n = 1), no incidence or prevalence data (n = 1), no treatment outcomes or treatment outcomes in fewer than two patients (n = 5), incidence or prevalence data from non-UK countries (n = 2)
- other reasons: same data reported elsewhere (n = 1), data included in existing systematic review (n = 1).

Quality of included studies

Incidence of GAI

No additional evidence was available on the incidence of GAI in the UK.

Effectiveness of treatments for GAI (Appendix 32)

Two studies provided information on the effectiveness of dietary interventions for GAI.

Bjugstad and colleagues[71] performed a forward, stepwise, multiple regression analysis to find predictors for outcome in 115 previously described patients with GAI using archival data from 42 published research articles. Non-biochemical variables included age at symptom onset, motor deficits, cortical atrophy, basal ganglia atrophy, enlargement of cerebrospinal fluid spaces, changes in white matter, treatment and clinical outcome (changes in motor behaviour). For this analysis, the authors combined data when one patient was a subject in two articles; therefore, no patients were double counted.

Most patients included in this study were started on a protein (lysine/tryptophan)-restricted diet supplemented with carnitine and/or riboflavin after the onset of symptoms. Of the 103 patients for whom treatment data were available, 15.5% had no treatment and 84.5% were treated. Treatment consisted of diet only (5.8%), carnitine or riboflavin supplement (32.0%) and carnitine and riboflavin supplement (46.6%). The authors reported that no statistical evaluations were performed to evaluate the benefits of presymptomatic treatment because only six patients were found to have been treated for GAI before the onset of symptoms with adequate biographical information.

The authors did not provide any information on age, gender characteristics, inclusion or exclusion criteria, quality of included studies, details of dietary compliance, duration of treatments or number of patients lost to follow-up.

Monavari and Naughten[69] evaluated the effects of treatment and clinical outcome in 12 patients with GAI by analysing retrospective clinical data. Patients in this study were diagnosed with GAI at the National Centre for Inherited Metabolic Disorders, The Children's Hospital, Ireland. Six patients were diagnosed as a result of family screening and six were diagnosed late after symptomatic presentation. Diagnoses were based on abnormal organic acids in urine by capillary GC/MS and clinical suspicion. The disorder was confirmed by enzyme assay of glutaryl-CoA dehydrogenase activity in cultured skin fibroblasts.

Management of the disorder was based on dietary regulation. The basic diet consisted of a synthetic protein drink (deficient in the amino acids lysine, hydroxylysine and tryptophan), natural (ordinary food) protein restriction but sufficient energy for growth, oral or intravenous supplementation with L-carnitine (100–200 mg/kg per day) and avoidance of catabolic state. The total protein intake ranged from 1.5 to 3 g/kg body weight per day (range of natural protein, 0.5–2 g/kg per day; synthetic protein, 0.5–2 g/kg per day; tryptophan, 5–21 mg/kg per day). Patients also received riboflavin; however, all were non-responsive.

During periods of acute illness or subtle metabolic decompensation, patients reduced natural protein intake (even stopping it for a short period of 24–48 hours, before reintroducing it gradually as tolerated clinically and biochemically), continued synthetic protein, increased energy intake by 20–50% by using oral and intravenous glucose, and doubled lipids and L-carnitine intake to 200 mg/kg per day.

The authors did not provide any information on age, gender characteristics, inclusion or exclusion criteria, details of dietary compliance or duration of treatments. Six patients were lost to follow-up (five patients died in the symptomatic group and one patient who was detected as result of family screening died).

The strength and quality of the evidence for the effectiveness of dietary interventions for GAI were grade III.[69,71]

Summary of results

Incidence of GAI

No additional evidence was available on the incidence of GAI in the UK. Owing to the lack of additional evidence, incidence data from the existing systematic reviews[1,2] have been reported.

Pollitt and colleagues[1] reported that there were no definitive data available on the incidence of GAI. Based on estimates from the number of cases clinically diagnosed in the UK (compared with similar disorders of established incidence), Pollitt and colleagues[1] estimated the expected incidence as two cases per 100,000 births. No incidence data were reported by Seymour and colleagues[2] for GAI.

Effectiveness of treatments for GAI

In the retrospective analysis of archival data from 42 published articles describing 115 patients with GAI, Bjugstad and colleagues[71] found that more than 50% of patients had an onset of symptoms before 8 months of age, and nearly all children in whom symptoms developed had the symptoms within the first 3 years of life. Only 8.7% of patients were reported as being asymptomatic, approximately 20% eventually improved to lead a normal life with normal or no residual consequence of GAI, and 21.6% of patients died.

A forward stepwise, multiple linear regression analysis showed that in patients who did not have a precipitating illness before the first appearance of motor symptoms, the age at onset was significantly associated with the severity of motor impairments and overall clinical outcome. Patients with symptom onset between 6 and 9 months had a much higher probability of a poor outcome, disease progression and death. In patients with gradually progressive symptoms (no precipitating illness), later onset of symptoms predicted better clinical outcome ($r = 51$, $p > 0.0004$). For patients who had a precipitating illness such as respiratory infection, vomiting or diarrhoea, no correlation was seen between the age at symptom onset and clinical outcome ($r = 0.003$, $p = 0.98$). Treatment given after the appearance of symptoms was not associated with a better clinical outcome or fewer motor deficits.

The authors reported that there was a lack of data to analyse the statistical benefit of treatment when it was given before symptoms occurred. Only six patients started treatment before any motor symptoms were present, all of whom have had a relatively normal development.

Monavari and Naughten[69] found that four of the six patients detected as result of family screening were developing normally, one patient had died and one had mild mental disability. The five remaining patients were aged between 3 and 9 years and were diagnosed with the disorder between 1 and 6 weeks. In these presymptomatic patients, supplementation with L-carnitine and intensive dietary treatment with aggressive emergency management resulted in a favourable outcome. Overall, the data showed that early diagnosis and aggressive treatment led to favourable outcomes with prevention of major neurological sequelae. In contrast, all of the six late-diagnosed symptomatic patients suffered from dyskinetic cerebral palsy and five patients died. The age of presentation in these six symptomatic patients ranged from 3 to 9 months; however, the diagnosis was deferred by 6–24 months owing to a delay in organic acid results (1–16 months) or a lack of awareness of the condition.

Discussion

No additional evidence was available on the incidence of GAI in the UK. Data reported in the Pollitt review[1] suggest that the incidence of GAI was two cases per 100,000 births in the UK. This estimate is based on the number of cases clinically diagnosed within the UK. The authors reported that this figure was a conservative estimate as the disease was widespread among a variety of racial groups and was probably underdiagnosed.

The evidence for the effectiveness of dietary treatments for GAI was limited to a few studies. Despite the poor study quality, the results showed that the age of onset of the disease could significantly predict the severity of motor deficits and overall clinical outcome. Dietary interventions given after the onset of GAI symptoms were not effective in reversing neurological damage[69] and were not associated with a better prognosis[71] in symptomatic patients. In contrast, dietary interventions offer some hope to presymptomatic patients in that early intensive management can alter the natural history of GAI; however, the threat of disability or death remains, and emergency management needs to be initiated with illness at all ages.[69]

Conclusion

Despite the lack of additional evidence on the incidence of GAI within the UK, the disorder is associated with increased morbidity and mortality and patients with GAI need to be identified as early as possible. Management of this disorder is less clearly defined; however, therapeutic strategies involving dietary restriction appeared to be effective in presymptomatic patients. Administration of L-carnitine to asymptomatic individuals with GAI appeared to offer considerable benefit by preventing the occurrence of neurological damage and allowing normal development. In symptomatic patients, the efficacy of the treatment had been extremely modest, with either no response or limited recovery of some functions. Data on the sensitivity and specificity of neonatal screening for glutaryl-CoA dehydrogenase deficiency using tandem MS are limited.

Pollitt and colleagues[1] concluded that GAI was relatively common and extremely serious, and there were indications that presymptomatic treatment could possibly prevent disease progression. Data on the sensitivity and specificity of neonatal screening using tandem MS were limited; however, several cases of GAI have been identified from neonatal dried blood-spot samples using tandem MS.

Seymour and colleagues[2] reported that GAI was associated with increased morbidity and mortality and that all patients with GAI have had a period before onset during which it is believed that intervention could have significantly improved outcome. These authors concluded that GAI comes close to fulfilling the necessary criteria for newborn screening, although there is a lack of information on the absolute incidence and natural history of the disorder. Prevention of the clinical manifestations of GAI (and of MCAD deficiency) alone could make newborn screening for these disorders cost-beneficial using tandem MS.

Multiple acyl-CoA dehydrogenase deficiency

Background

Multiple acyl-CoA dehydrogenase deficiencies (glutaric aciduria type II or GAII) cover a group of conditions resulting from the impairment of the process of dehydrogenation of many different acyl-CoA esters. This relatively rare condition has a wide range of clinical severity, with the most severe effects being multiple congenital abnormalities that lead to death in the neonatal period. Less severe presentations are identified in the first week of life and involve profound metabolic decompensation and acidosis.

Untreated cases surviving the neonatal period frequently die in the course of recurrent attacks. Milder cases may present later with acute Reye-like symptoms similar to MCAD deficiency or follow a more chronic pattern with poor feeding, failure to thrive and hypotonia.

Quantity and quality of research available

Quantity of research available

The yield of literature from the systematic searches generated 69 additional publications on the incidence, prevalence and effectiveness of treatment outcomes for GAII. Full copies of four papers were retrieved for further evaluation.

Of these four papers, none was judged to be relevant to the update review. These studies were excluded from the update review (details of references are provided in Appendix 33), for the following reasons:

- study design: screening ($n = 1$)
- outcome measures: no incidence or prevalence data ($n = 1$), incidence or prevalence data from non-UK country ($n = 1$)

- other reasons: same results reported elsewhere ($n = 1$).

Quality of included studies

Incidence of GAII

No additional evidence was available on the incidence of GAII in the UK.

Effectiveness of treatments for GAII

No additional evidence was available on the effectiveness of treatments for GAII.

Summary of results

Incidence of GAII

No current evidence was available on the incidence of GAII in the UK. Owing to the lack of additional evidence, incidence data from the existing systematic reviews[1,2] have been reported.

Pollitt and colleagues[1] and Seymour and colleagues[2] reported that this disorder was relatively rare and there were no reliable data available. Based on estimates from the number of cases clinically diagnosed in the UK (compared with similar disorders of established incidence), Pollitt and colleagues[1] estimated the expected incidence as two cases per 100,000 births.

Effectiveness of treatments for GAII

No additional evidence was available on the effectiveness of treatments for GAII. Pollitt and colleagues[1] and Seymour and colleagues[2] reported that there were no effective therapeutic interventions for the severe forms of GAII. Treatments for the milder variants included a moderate reduction in dietary protein and fat intake, and increased carbohydrate intake. Some patients with mild variants respond favourably to high-dosage riboflavin and L-carnitine supplementation; however, action must be taken to reduce catabolism during periods of illness.[1,2]

Discussion

No additional evidence was available on the UK incidence and effectiveness of treatments for GAII. Information for these disorders was solely derived from the two existing systematic reviews.[1,2]

Conclusion

Despite the lack of additional evidence on the incidence of GAII in the UK, data reported in the Pollitt and Seymour reviews[1,2] suggest that the incidence of GAII was relatively low. Pollitt and colleagues[1] estimated the expected incidence as two cases per 100,000 births. Evidence is very limited regarding the sensitivity and specificity of neonatal screening for GAII using tandem MS.

Pollitt and colleagues[1] concluded that there is little benefit in diagnosing early presenting cases, except for the provision of genetic information. In contrast, moderately severe and milder cases may benefit considerably from early diagnosis. Severe and moderate forms of GAII are detectable by tandem MS; however, there is no information as to the sensitivity for very mild variants.

Seymour and colleagues[2] concluded that specific screening for GAII should not be considered owing to the wide range of presentations, unknown incidence (relatively rare) and relatively poor outcome of treatment. However, it is probable that affected individuals would be detected during newborn screening using tandem MS.

Chapter 7

Economics of neonatal metabolic screening

Introduction

Economic issues in health technology assessment

Although screening represents a distinctive form of health technology, the fundamental economic need to evaluate the efficiency of the intervention remains. The distinctiveness of screening relates to the specific identification of appropriate benefits and costs that should be included. In this context, three main considerations should be explicitly addressed: clarity about the viewpoint or perspective taken, how efficiency is defined and measured, and the role of equity.

Viewpoint refers to the breadth of perspective taken on the evaluation of a screening programme. Although different viewpoints may be valid, economists tend to prefer where possible to assess health technologies from a broad 'societal' perspective. This is justified on the grounds that it is appropriate to be concerned about how investments in health technologies maximise overall levels of welfare available from society's scarce resources.

The viewpoint adopted has significant implications for the range and types of costs and benefits that should be included within an evaluation. For a screening intervention, for example, costs should not only comprise those resources directly consumed by the provision of a screening programme, but also include any subsequent health and social care costs that arise from presymptomatic or symptomatic detection. This may comprise future treatment costs incurred or avoided as a consequence of screening. Other healthcare and social care costs related to any disabilities or impairments that arise from those affected or treated, with and without the presence of screening, should also be considered.

Other significant resource consequences affected by screening are (net) future changes to productivity and other indirect costs. These may arise either from the impact or avoidance of future mortality and morbidity on affected individuals, or from changes to the burden of support on informal carers. However, ethical concerns over the relevance of identifying the future productive capacity of affected individuals, as well as significant methodological differences over the quantification of indirect costs, frequently render their inclusion problematic. Good practice dictates that indirect costs, if included, should be identified and reported separately; and this should include explicit reference to the basis on which future productivity costs have been calculated.

Inappropriate perspective can lead to a narrow or unsuitable definition of costs and benefits to be included within an evaluation. In the context of screening, one consequence of this has been "the choice of overly simplistic, inappropriate maximands", such as "cost per case detected".[72] Yet false-positive and false-negative results within a screening programme are also important. Both are likely to impose significant costs and may substantially affect the net benefits attributable to a screening intervention. These issues are relevant to the overall measurement of programme efficiency.

Fundamentally, efficiency is about comparing benefits obtained against the resources consumed. However, economists frequently distinguish between two types of efficiency: technical and allocative. 'Technical efficiency' is conventionally defined in terms of maximising the benefits (outputs) obtained from a given programme from the resources used; or, alternatively, minimising the resources depleted for a given level of benefits (outputs). Technical efficiency is primarily about how services are provided and aims to optimise the use of resource inputs (capital, materials and labour) for a given intervention or health technology. Thus, technical efficiency would seek to identify the best combination of inputs (given the relative price of those inputs) for delivering, for example, a neonatal screening programme for phenylketonuria.

Allocative efficiency recognises that healthcare inputs can be used for many different purposes and seeks to maximise total health benefits available from the use of those resources. This is a wider notion of efficiency: instead of looking at how best to provide a particular service or health technology, it examines what combination and level of health technologies should be provided from the available healthcare budget.

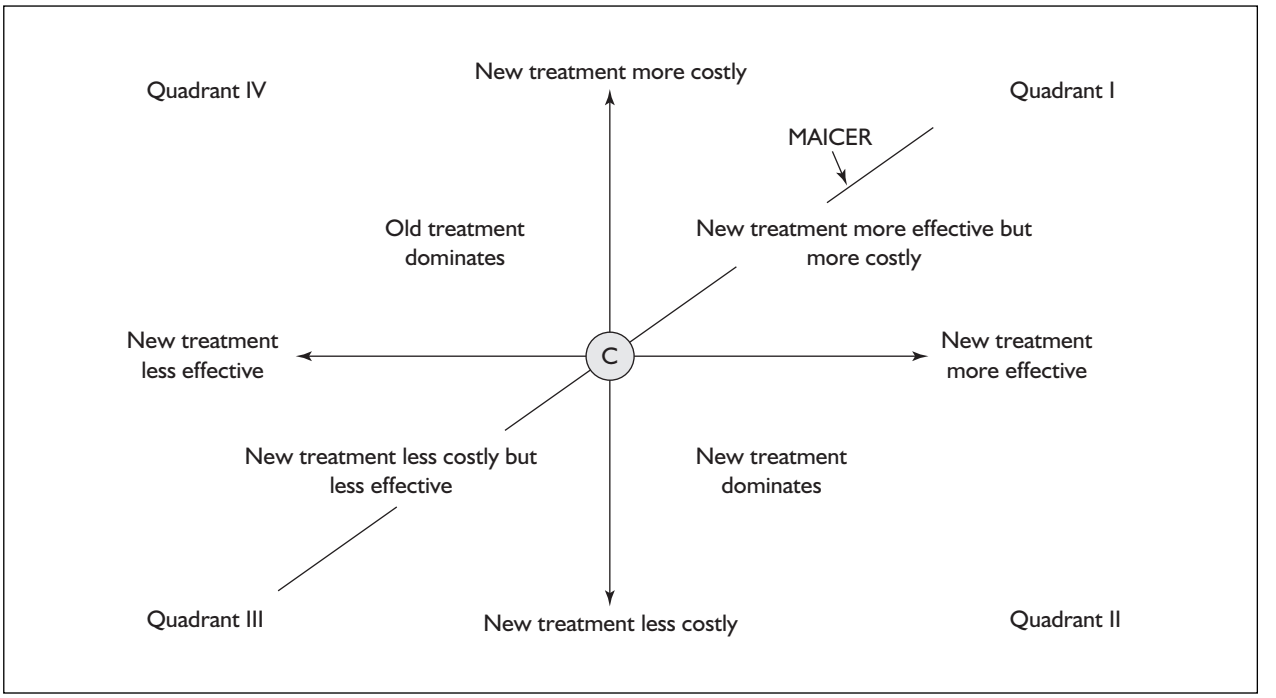

FIGURE 1 *The cost-effectiveness plane (example). MAICER, maximum acceptable incremental cost-effectiveness ratio*

The results of an economic evaluation of new technologies can be represented on a cost-effectiveness plane.[73,74] Overall, the results of an evaluation may produce four different outcomes based on the calculated incremental cost-effectiveness ratio (ICER). This is illustrated in *Figure 1*. Note that the axes in the diagram are presented in terms of incremental costs and effects. It should be clear that only two of the four situations produce dominant strategies (quadrants II and IV); the other possible outcomes involve results that are not unequivocal and will require decision-makers to trade off the relative magnitude of cost–effect differences produced.

Figure 1 also illustrates how cost-effectiveness analysis may be of limited value to decision-makers where the choice of strategy suggested by the incremental cost-effectiveness ratio does not clearly dominate. For example, a new technology may produce a lower cost-effectiveness ratio for a given set of objectives, but in total actually consumes more resources (quadrant I, in *Figure 1*). Selection of a new technology that costs more in total will require the diversion of resources from alternative healthcare interventions. This will impose opportunity costs in the form of foregone benefits owing to the reallocation of resources. Although a new intervention may be technically more efficient, there may be issues of allocative efficiency that need to be considered before a final decision can be made.[75]

Ideally, the evaluation of a new technology should be expressed in costs and benefits that also facilitate the consideration of allocative efficiency. The question, therefore, is not just about which technology provides the most cost-effective way of delivering a particular screening programme, but also about whether the screening programme per se represents a good investment of healthcare resources relative to the alternative uses to which these resources could be put. In this context, judgements about the relative value of pursuing new healthcare technologies or forms of service delivery may be made in terms of an acceptable 'threshold' of incremental cost–benefit. This threshold is usually set in terms of a maximum acceptable incremental cost-effectiveness ratio (MAICER) and, increasingly, this is being defined in terms of an acceptable cost per quality-adjusted life-year (QALY) ratio (e.g. a MAICER of £20,000, £30,000 or £40,000). This simplifies the decision rule: any cost-effectiveness ratio that is less than the MAICER is worth pursuing (i.e. any ratio that falls to the right of the diagonal line in *Figure 1*).

It is important to understand that this threshold will not be fixed but may vary according to society's willingness to pay for different health benefits that accrue in different circumstances. For example, society may have different cost-effectiveness thresholds for health benefits that affect particular groups, such as those that affect children, those already disadvantaged by chronic

disabilities or interventions targeted at correcting health inequalities.

Although economic evaluations have traditionally been designed to provide decision-makers with information on the relative efficiency of different options for achieving certain objectives, the role of equity should not be overlooked. Equity concerns the 'fair' distribution of available costs and benefits from health technologies to different groups within society. The complexities of definition and measurement are not relevant here; our concern is to indicate that equity provides an important additional dimension to that of efficiency, which decision-makers need to acknowledge.

Equity considerations are especially important where the distribution of benefits affects various social groups differently. In circumstances where particular disadvantaged or ethnic sections of the population are more susceptible to conditions for which screening offers the opportunity for early treatment, equity may override a strict adherence to relative efficiency expressed in terms of cost per unit of benefit calculated on average measures of effectiveness.

Methods for economic analysis

Evaluation methods

Different techniques are available with which to evaluate the efficiency of a health technology. The choice of method depends on the objectives of the study, the viewpoint taken and the types of costs and benefits that need to be quantified. A useful summary description of the various types of economic evaluation that can be used to assess neonatal screening was provided in a previous HTA report.[1]

Cost minimisation analysis would be unsuitable for the evaluation of whether screening should or should not be performed because the outcomes (health benefits) of the two strategies will not be equal. Cost minimisation may be appropriate, however, as an assessment of whether an existing screening technology can be replaced by a newer technology, if there is good evidence that the outcomes achieved by the two alternative technologies are identical. Thus, if the outcomes obtained using tandem MS for the screening of PKU are the same as those achieved by existing technologies (Guthrie, fluorometry and chromatography), then the evaluation can be reduced to a consideration of which produces the least-cost method of screening. Although there are very few circumstances where the outcomes from applying different technologies are identical, the use of cost-minimisation analysis might be a reasonable conservative approach where there is evidence that the new technology is likely to perform better (e.g. has a higher sensitivity and specificity as a screening tool).

Cost–benefit analysis would be particularly appropriate for the evaluation of screening technologies because of its ability to capture all costs and benefits in terms that permit direct comparison. This could include the measurement of a number of important non-resource and externality effects that are recognised as important consequences of screening (including the reassurance or anxiety caused by screening) as well as the valuation of any (net) health gain.

Cost-effectiveness analysis (CEA) is the technique most commonly applied because of the relative ease with which the benefits can be measured. This is a technique that expresses benefits of healthcare interventions in terms of 'natural units'; units that are often readily defined by the type of intervention provided. Examples include symptom-free days, sight-years saved, true cases detected or life-years saved. These units often have intuitive appeal or understanding to those involved in a particular technology or service provided. However, conventional CEA fails to measure the complete health benefits or gain available from an intervention because it does not explicitly capture the trade-offs between additional years of life and the quality of life provided. It is also a technique that restricts comparisons between technologies that can only be defined in terms of the same natural units of benefit. Although CEA may address issues of technical efficiency, decisions about allocative efficiency are constrained by the absence of a common denominator that allows health gains from different types of technology to be compared.

Cost–utility analysis (CUA) does permit comparisons across different technologies through the measurement of benefits using an index that combines both the quantity and quality of life effects of an intervention (a QALY). Future life-years obtained from a health intervention are weighted using an index that adjusts a year of life according to the assessed quality of that life. These (QALY) weights are normally obtained using a number of generic health-related quality of life (HRQoL) instruments designed to elicit preferences for different health states. QALY weights can be obtained in different ways. There are examples where those providing an intervention estimate

the health state values attached to a particular outcome obtained, with or without treatment. The inherent weakness in this approach is that those who provide an intervention may have a biased understanding of the health impact of a particular condition and of the benefits of treatment. The more conventional approach used as part of a study trial is to obtain health state valuation from those individuals directly affected.

CUA would provide a more appropriate basis for the evaluation of the benefits of neonatal screening for particular metabolic disorders than CEA, which failed to capture the main quality of life effects related to the provision or absence of screening. Again, however, the literature surveyed showed no recent examples of cost–utility studies in respect of neonatal screening for inborn errors of metabolism. The authors were aware that a number of economic evaluations on the use of tandem MS for screening of inborn errors of metabolism were underway, but these studies were not published in time to be included within this review.

Uncertainty and the interpretation of results

No matter what type of economic evaluation technique is used, the overall robustness of the results obtained should be tested to establish the probability that the right option has been identified. This is necessary because of the inherent uncertainty in any estimates of the costs and effects derived from an economic evaluation. The standard way of dealing with uncertainty in the main quantifiable elements of a study (the parameters) has been the application of sensitivity analysis. Parameter uncertainty will arise because some of the study elements cannot be observed directly. Moreover, in the absence of an infinite sample, actual study parameters will always contain some degree of statistical variability.

One-way (univariate) sensitivity analysis examines the impact of each parameter on the overall study results by changing their values one at a time. In a univariate approach a specific parameter's values are varied by some plausible amount and the cost-effectiveness ratio is then recalculated. From this it is possible to identify which variables, and their associated uncertainty, have the greatest individual effect on relative cost-effectiveness. However, 'cost' and 'effect' estimates in a study are not normally calculated using single parameters, but are based on a series of variables, each of which have uncertainty, connected both to their values and to their interrelationships.[76] Therefore, although useful by itself, univariate sensitivity analysis is recognised as inadequate; examining the effects of changing one parameter at a time provides an incomplete assessment of the overall uncertainty within study results.

As an alternative, multiway analysis allows for more than one parameter at a time to be varied. The choice of which variables to include within a multiway analysis may be made by an initial univariate procedure which identified those parameters that are likely to have the most significant impact on the study results. As a relatively straightforward special case of multiway analysis, all of the most favourable or most unfavourable parameter values can be combined simultaneously to produce a 'best' and 'worst' case scenario, respectively.

One frequently observed weakness of sensitivity analysis in general, however, has been the often arbitrary way in which parameters have been varied; for example, a doubling or halving of parameter values as a way of estimating best and worst case scenarios. More appropriate methods for capturing a plausible range have been suggested.[77] These methods include systematic reviews of the literature, reference to expert opinion, and use of statistical procedures to calculate confidence intervals around parameter values.

Partly as a response to these concerns, procedures have been developed to estimate uncertainty using statistical analysis of the sampling variation of the parameter used within a study. Different approaches are possible. Of particular interest methodologically has been the refinement of methods for calculating confidence intervals around the ICER. These methods can be complex technically and most are grounded within the classical (or frequentist) tradition of statistical theory, which calculates confidence limits based only on the observed data obtained within a study. To capture the inherent stochastic uncertainty of the results, calculation is often based on methods that require repeated sampling from within the plausible range of parameter values. For trial-based evaluations these can be generated using bootstrapping techniques. The results are usually presented diagrammatically, depending on the procedures used, in terms of a box or an ellipsoid showing a confidence region around the ICER. In appearance, this is similar to a distribution of values within a quadrant of the cost-effectiveness plane (*Figure 2*).

More recently, attention has shifted to the use of an alternative Bayesian framework. One distinctive feature of Bayesian approach is that inference is

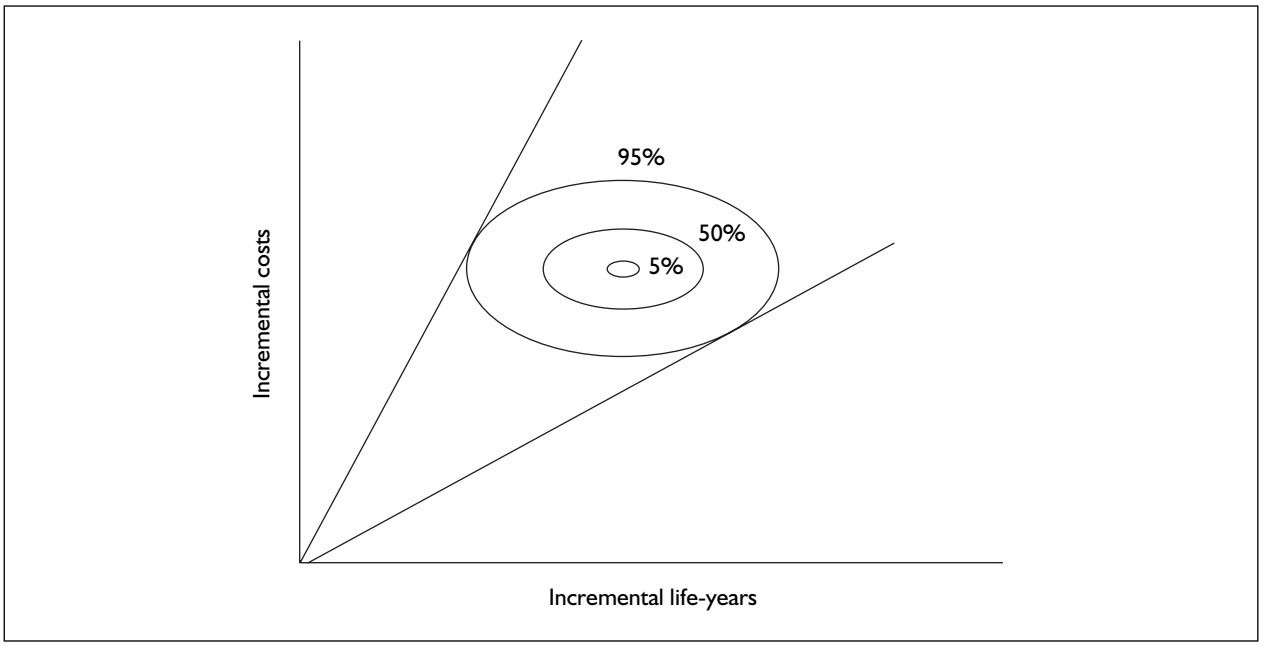

FIGURE 2 *Confidence ellipse on a cost-effectiveness plane (example)*

based not only on the observed data, but also on a probability distribution that will characterise what is also known about specific parameters. This may include prior information not captured as part of any study data collection.[76] Information and assumptions about the prior distribution of parameter values are critically important to Bayesian methods. The Bayesian approach, which allows the synthesis of information from various sources, may be particularly suited to the evaluation of health technologies (such as neonatal screening for rare inborn metabolic disorders) that are not susceptible to conventional RCTs.[78] The Bayesian equivalent of the confidence region is the credible region.

Nevertheless, calculation of confidence (or credible) regions can be problematic. In addition to differences that may emerge according to which method of calculation is used, more fundamental problems may arise when the cost and effect differences are not confined to a single quadrant of the cost-effectiveness plane and so produce indeterminate results.

Particular difficulties arise when the calculated ICERs are negative. Briggs and Gray provide a simple example to illustrate the point:

> "If an intervention is associated with an increase in life-years of 0.25 years and a cost saving of £1000 it will have an ICER of –£4000 per life-year gained. It is clear that an alternative intervention which generates the same cost savings but generates a health benefit of 2 years will be preferred, but it will have an ICER of –£500 per life-year, which is greater than –£4000."[74]

One resolution of this problem is based on the notion of a threshold value as the appropriate decision rule, discussed earlier. Any calculated ICER less than the accepted ceiling ratio (the diagonal MAICER line in *Figure 1*) should be accepted. As stochastic analysis may produce ICER values across the cost-effectiveness plane, the overall uncertainty can be summarised using the proportion of ICER values that fall to the right of the threshold value. Since the threshold value may not actually be known or defined, the actual ceiling ratio (MAICER) can be varied to indicate how the relative cost-effectiveness of the new technology might compare with different threshold values. The resulting set of calculations can then be used to describe a cost-effectiveness acceptability curve (CEAcc).[79]

The CEAcc plots the proportion of cost-effect pairs generated by the stochastic process that indicates when the new technology is optimal relative to the threshold (*Figure 3*). A further advantage of the CEAcc approach is that the combination of cost and effect pairs with established threshold values (MAICERs) allows the net benefit of any new technology to be estimated.

Net benefit refers to the difference in the expected pay-off from a new intervention (expressed in

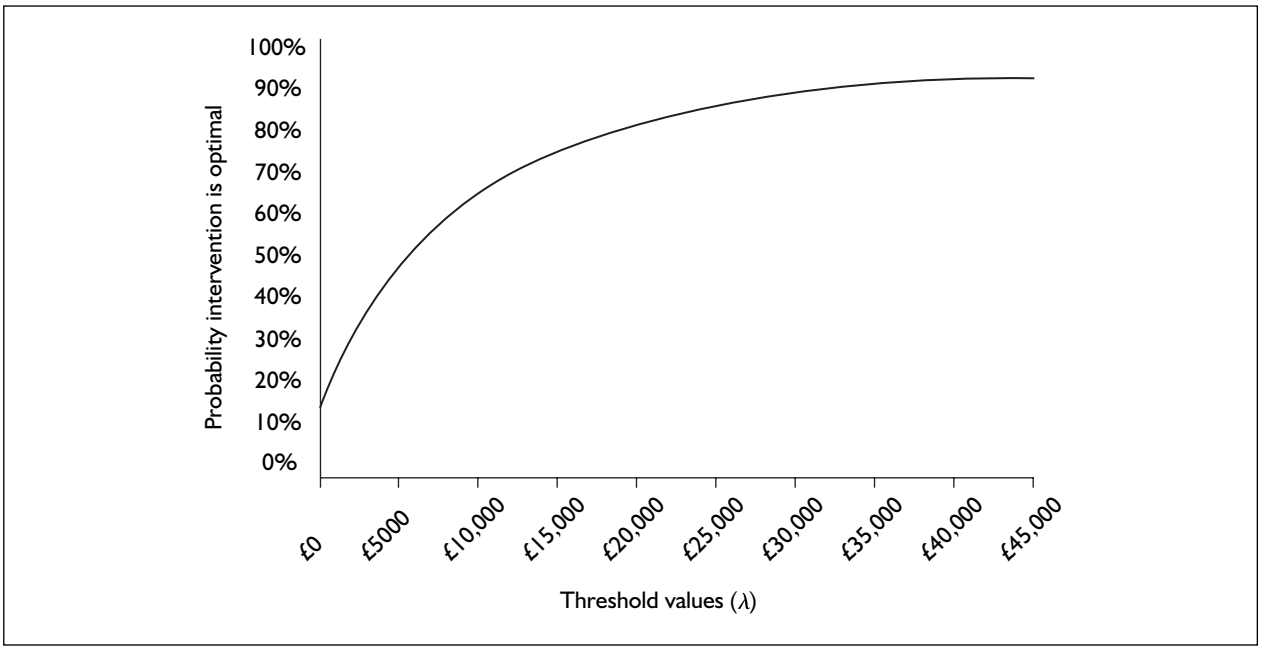

FIGURE 3 *Cost-effectiveness acceptability curve (CEAcc)*

monetary terms) and the threshold value (λ) set by decision-makers about what constitutes an appropriate return on healthcare technology. This threshold (λ) is usually set according to some agreed notion of what represents society's willingness to pay for an additional unit of health benefit (i.e. its maximum acceptable cost-effectiveness ratio).

Net benefit can easily be operationalised by expressing health benefits in monetary terms using the implicit valuation provided by the threshold value. A simple example should serve to illustrate the concept. First, imagine that society has set its threshold value (λ) at £30,000 per QALY and the evaluation compares two treatment options for a specific condition. Further, assume that the QALYs gained from our new intervention (T1) are five and the QALYs obtained from the existing intervention (T0) are three, and programme costs are estimated at £100,000 and £55,000, respectively. Incremental net benefit (INB) can then be expressed as:

$$\begin{aligned} \text{INB} &= \lambda * (E_{T1} - E_{T0}) - (C_{T1} - C_{T0}) \\ &= £30{,}000 * (2) - (£45{,}000) \\ &= £15{,}000 \end{aligned}$$

This approach has the advantage of providing a simplified decision rule: in general, a new technology should be accepted when the INB is greater than zero (i.e. if INB > 0, then fund). Furthermore, higher net benefit values are always better; the potential indeterminacy of some ICER values is avoided.

Moreover, as indicated above, different thresholds (λ) can then be set and stochastic analysis of cost and effect pairs generated to calculate the probability that a new intervention will produce an INB greater than zero at each threshold value. The results can then be presented as a CEAcc.

Economic modelling

So far, this summary of the techniques used by health economists has been presented in terms of standard trial-based evaluations. The same stochastic properties and analytical considerations can be applied to economic models. In this case, however, repeat sampling using bootstrap methods is replaced by repeated iterations using Monte Carlo simulation.

Economic modelling is now recognised to be a valuable technique in the evaluation of new healthcare technologies. The advantages of a model are that they can be formulated using a range of evidence sources and do not require a relatively expensive piece of primary research (e.g. through prospective studies such as RCTs). Modelling can rely on the synthesis of evidence from a variety of secondary sources. These may include published studies in peer-reviewed journals, reports and technical documents, as well as extracts from administrative datasets and information derived from expert panels.

Modelling techniques also allow the process of evaluation to be extended beyond what has been directly observed.

Although all modelling, by definition, provides a simplification of reality, its effective use does presuppose that there is sufficient information already available that is capable of providing a reasonable heuristic representation of that reality. This means that information is available which allows estimation of the values associated with the main parameters required for construction of the model. There is a manifest compatibility between modelling and a Bayesian methodology that allows a broader integration of parameter and probability data from different sources.

Whatever evaluation techniques are used, the extent of uncertainty surrounding the parameter estimates directly affects the overall robustness of the results obtained. Another advantage of the modelling process is that the effects of uncertainty on the relative cost-effectiveness of any new technology can be explored in more detail. This is the essence of sensitivity analysis described earlier; but the methods available allow for greater sophistication.

Two main types of modelling approach can be distinguished. Deterministic modelling is primarily a method that relies on using a combination of fixed or mean values within the modelling scenario. In contrast, probabilistic modelling uses the whole distribution of parameter values that are obtained, or can be realistically defined, and combines them in random ways using iterative techniques such as Monte Carlo simulation. The use of such techniques within a Bayesian framework means that economic modelling is particularly appropriate to a probabilistic estimation of uncertainty surrounding a new technology. Increasingly, this aspect of economic modelling is undertaken using probabilistic methods within a formal value of information (VOI) framework.

The value of information

A formal VOI analysis seeks to estimate the expected costs associated with imperfect information when selecting between alternative courses of action. This can be characterised as uncertainty. Uncertainty affects the extent to which we can be confident that our current information allows us to identify what is the underlying best option and that the estimated benefits would be obtained in practice should the new strategy be adopted. Uncertainty means that there may be circumstances when the net benefit of the alternative strategy would have been better.

The value of additional information on a specific question, therefore, depends on the extent to which this additional information will reduce the uncertainty about the selection of alternative strategies. Information has value because it reduces uncertainty and thereby the potential costs (the opportunity loss or benefits forgone) of selecting a particular option as optimal when it is not.

The expected costs of uncertainty can also be seen as equivalent to the expected value of perfect information (EVPI). Decision-makers with perfect information could always be entirely sure that the right (optimal) choice had been made.

Estimation of the EVPI can then be used to determine whether the costs of any additional research designed to reduce uncertainty are worthwhile.[80] If the EVPI is greater than the expected costs of acquiring the extra information, then it is cost-effective to carry out additional research. This is based on a simple premise: if the benefits obtained from additional research, in terms of reduced uncertainty, exceed the costs, then the research should be conducted.

A distinct advantage of an EVPI analysis is that not only is it possible to estimate the expected costs of uncertainty across all parameters, but the expected costs associated with individual parameters can also be estimated separately.[81] This means that future research, if required, can then be focused on those inputs that are likely to have the greatest impact on the decision question.

Although CEA and VOI methods are complementary, and so frequently used together, they are nevertheless distinct. Conventional CEA is designed to establish the relative value of adopting different health technologies by assessing the incremental costs and benefits associated with alternative strategies or interventions. Given uncertainty around the CEA results obtained, VOI analysis then estimates what would be the value of additional research in order to reduce this uncertainty.

Conclusion

This brief survey of the main conceptual tools and methods now used by health economists in the evaluation of healthcare technologies is of direct relevance to the use of economic modelling and the interpretation of the model results presented later in this report.

Identifying and quantifying the costs and benefits of neonatal screening for inborn errors of metabolism

Economic efficiency requires a consideration of both the costs and benefits of a healthcare programme. The following two sections briefly set out the main categories of each that can be considered within an economic evaluation. Precisely what type of costs and benefits will be included depends on the viewpoint taken.

Costs

Specimen collection costs

For newborn screening of inborn metabolic conditions, the specimens collected are blood spots taken from the neonate and impregnated on to special Guthrie cards. These cards are then sent immediately to laboratories for analysis. In the UK, midwives (in some areas health visitors) collect initial blood samples as part of their statutory visits to follow-up infants carried out within 6–14 days after birth.

As the blood specimens are collected as part of a routine visit designed for assessment as part of a child health surveillance programme, only a proportion of the cost of the midwives' contact should be attributed to specimen collection. Only when repeat samples of blood are required would a health visitor make a further special visit to collect a specimen.

One of the technical advantages of tandem MS technology is that additional metabolic disorders can be screened for using the same blood specimen collected and analysed for the presence of PKU. Extending neonatal screening to include the other metabolic disorders reviewed here, therefore, would require no additional sample collections. However, extension of neonatal screening would incur some small additional cost for initial sample collection because extra time would be required by midwives to explain what additional conditions are being screened for and to obtain consent; this should also include provision of some additional informative materials to parents. Many parents, however, are already informed about the issue of neonatal screening in advance through attendance at parentcraft classes run by midwives.

Laboratory costs

Once the specimens have been collected they are dispatched for analysis at designated neonatal screening laboratories. The costs of screening in the laboratory setting involve three distinct elements: labour, capital, and consumables.

Staff inputs are required to organise and document the receipt of specimens to be examined. Specimens also need to be prepared (blood spots punched into microtitre plates with appropriate solvents) and loaded for analysis, according to the screening method used. Increasingly, these elements have become mechanised to facilitate the handling of large volumes. Staff are also required to interpret the results of the laboratory analysis and initiate any subsequent actions (e.g. reporting, repeat specimen requests, confirmation protocols).

Newer technologies such as tandem MS allow for the automation of results profiling and interpretation based on computer algorithms.

The equipment required to analyse the blood specimens depends on the technology used. In the UK, PKU screening may be performed using a Guthrie, fluorometric or chromatographic technique. Each of these involves different forms of equipment with different degrees of capital outlay. Tandem MS is a relatively capital-intensive technique, with a large initial outlay for purchase of the necessary equipment. Capital costs are usually fixed over a large range of processing outputs and provide benefits over future years (depending on the life expectancy of the equipment).

Consumables required for laboratory screening will also vary somewhat according to the method used. Items here include the use of solvents and suitable analytical containers (e.g. microtitre plates). These are variable costs that largely increase in direct proportion to the volumes of specimens processed. Maintenance costs for each of these technologies also need to be included as an annual cost. To this can be added an annual charge for facility overheads (usually apportioned on the basis of area of accommodation used).

For the purposes of this review, and particularly the economic modelling presented later, the combination of Guthrie, fluorometric and chromatographic methods is regarded as the existing technology. In practice, it is recognised that a number of laboratories in the UK already use tandem MS to screen for PKU.[82]

Confirmation costs

After an initial test has indicated a positive result, it is necessary that this is confirmed before a definitive diagnosis is given and treatment initiated. Uncertainty surrounding the initial test result may occur for a number of reasons. Test

results may be within an equivocal region of the defined cut-off values used to establish the presence of a disease. In these circumstances additional tests will be undertaken to facilitate the correct diagnosis of the underlying condition. Moreover, for every true-positive case detected there will also be a number of false-positive results, depending on the specificity of the screening test. It is extremely important to eliminate these false positives as quickly as possible so that an incorrect diagnosis is not made and to reduce any possible anxiety caused to those affected.

Confirmation protocols will vary according to the types of metabolic conditions found. For PKU, confirmation may require additional biochemical assays (of biopterins and dihydropteridine reductase activity). For MCAD deficiency there may be a repeat blood-spot acylcarnitine profile and urinary organic acid studies; supplemented in a small proportion of cases by genetic mutation analysis and cultured skin fibroblasts, depending on how equivocal the other results are. There will be a broadly similar procedure for other organic acid disorders; for many of the non-fatty acid disorders a urine sample would normally be sufficient to confirm diagnosis. In addition to the affected individuals, siblings are often routinely tested when a positive diagnosis is confirmed.

Once a positive test has been confirmed, additional input from senior laboratory staff will be required to organise an appropriate referral. This often requires time on the telephone providing advice, arranging referrals and writing letters. This may be described as referral and advice.

Note that confirmation costs will vary depending on whether a particular condition is screened for or not. With screening there will be a proportion of false-positive cases that will need to be eliminated. False positives involve additional laboratory tests and specimen collection (e.g. urine for analysis). Moreover, for some disorders (e.g. MCAD deficiency) screening will detect a proportion of cases that would previously have remained asymptomatic. The cost implications of any systematic differences of this type should be included within the evaluation of any extension of neonatal screening to other disorders.

Acute episode costs related to symptomatic diagnosis

In the absence of a neonatal screening programme a high proportion of most of the inherited metabolic disorders will eventually present as symptomatic cases. Symptomatic detection is normally precipitated by an acute period of hospital care. Some of these acute hospitalisations will be serious and may require periods of intensive care (e.g. admission to paediatric critical care units).

There may be some limited circumstances when this difference in acute hospitalisations might not occur. This could include occasional presentation before screening results are available, or alternatively, some cases may also be detected presymptomatically without screening through the testing of siblings.

However, in the majority of cases it is reasonable to assume that there will be at least one acute episode that precipitates symptomatic diagnosis. As a corollary, it can also be assumed that most of these specific diagnostic episodes would be avoided by presymptomatic detection through screening. For MCAD deficiency there is evidence in the literature to indicate that many symptomatic patients have undergone a number of acute episodes before a correct diagnosis is obtained. How this issue is dealt with will be considered later as part of the description of the economic modelling assumptions.

Treatment costs

Once a definitive diagnosis has been given many of the metabolic disorders require some form of treatment intervention. For most of the conditions reviewed here this consists of some form of dietary intervention with advice on how to manage the condition to prevent subsequent episodes of decompensation.

Details of the treatment strategies applied to specific conditions, and the estimated costs of each, used in the economic model are set out in Appendices 34–36. They are the same as those used within the economic model produced as part of the Pollitt report,[1] updated to reflect current (2001) prices.

Treatment for each condition is normally initiated as soon as a diagnosis is made. For presymptomatic diagnosis (through neonatal screening), treatment and management will begin within the first few weeks after birth. In the absence of screening, treatment and active management will not begin until diagnosis occurs through acute presentation. As treatment strategies are effectively the same whether diagnosis is obtained as a neonate or through later presentation, the incremental costs of treatment relate mainly to the average period

that intervenes between presymptomatic and symptomatic detection.

It is recognised that the treatment profiles described in Appendices 34–36 may not now completely reflect current practice for certain disorders. In the USA L-carnitine supplementation is often routinely prescribed for the treatment of MCAD deficiency cases; and this therapeutic regimen is now being applied in the UK (although it is not standard practice). For some of the more severe conditions (e.g. methylmalonic and proprionic acidaemia, MSUD and some of the urea cycle disorders), there has been an increasing trend towards the use of liver transplantation (currently standard treatment for tyrosinaemia type I). Outcomes from liver transplantation have improved in recent years and these procedures will incur major additional early treatment costs. Clearly, this trend would be attended by a notable increase in costs for the treatment of inborn errors of metabolism and this would be an economic issue of real interest to the NHS as a whole. However, it seems likely the use of liver transplantation would not depend on the existence of a screening programme per se, and transplantation will only be used if judged appropriate, irrespective of whether the individual is diagnosed presymptomatically or symptomatically. As the question of relative cost-effectiveness of extended screening depends more upon incremental treatment costs, the increasing use of relatively expensive procedures such as liver transplantation will only be relevant where there would be significant systematic differences in the use of these newer techniques between screened and unscreened cohorts. (It is possible, for example, that some systematic differences for liver transplantation would arise from a reduction in the number of early deaths and because presymptomatic diagnosis would permit earlier intervention, before significant neurological damage has occurred. This represents an additional uncertainty in the current evaluation of these conditions.)

Similar arguments also apply to any significant changes in treatments applied to the emergency management of any acute episodes of recurrent illness, provided that there is no systematic difference in the number of acute episodes after diagnosis with and without the existence of a dedicated neonatal screening programme.

The validity of using specific treatment strategies, and the cost estimates for each, set out in Appendices 34–36, will be considered in more detail in those sections explaining the assumptions and prior distributions used within elements of the economic model.

Note that, as a proportion of cases for some conditions may always remain asymptomatic (e.g. MCAD deficiency), it should not be assumed that the numbers treated without a screening programme will be the same as those treated with a screening programme. An economic evaluation should reflect these differences by excluding the estimated number of asymptomatic cases from the future treatment costs incurred by the non-screened cohort.

Future healthcare and social care costs related to disabilities and impairments caused by acute presentation

Future episodes of decompensation will tend to occur in cases detected by neonatal screening and those detected symptomatically (depending on severity of the condition, compliance with treatment, events precipitated by general illness, etc.).

As indicated above, a significant difference between presymptomatic and symptomatic diagnosis, however, is that the latter may be attended by a series of serious acute episodes of illness before a definitive diagnosis is made and treatment is initiated. For some of the metabolic disorders considered here these episodes may not only be life threatening but also lead to major residual disabilities and impairments in a significant proportion of cases. These disabilities and impairments, although relatively small in number, will incur future healthcare and other social care costs that could have been avoided through presymptomatic detection and early treatment; for example, additional consultations with occupational therapist, physiotherapists and speech therapist; orthotic and orthopaedic procedures; primary care and community support; home adaptations; residential care; and special educational needs and support.

As with treatment costs, however, it is important to identify the additional costs related to any physical disabilities or cognitive impairments caused by metabolic disorders. For example, the estimated costs of any special educational support during childhood and adolescence should subtract the normal educational costs that would have arisen in the absence of disability. This focus on incremental cost should also be applied to all future healthcare and social care costs that may be incurred.

Indirect costs
Future indirect costs associated with those severely disabled or impaired by metabolic disorders are likely to be significant. Although the number of families involved will be small, the volume of productivity losses and informal care costs imposed by the extra care needs of affected children can be substantial.[83]

As indicated earlier, however, the inclusion of indirect costs – especially future productivity gains or losses – may be problematic given some of the methodological uncertainties around their inclusion and valuation. There may also be the more practical difficulty of reliably estimating these future costs in the absence of more extensive study evidence.

Benefits

Various health benefits can be considered as appropriate to the evaluation of a neonatal screening programme.

True cases detected
An ability to detect those actually with a specific disorder (the number of true positives) is probably the first and most transparent requirement of any screening technology. But, although this is a fairly obvious and important consideration, measuring the benefits of a screening programme entirely on the basis of the number of true cases detected is inadequate. First, detecting the presence of a disorder may be of limited value without the opportunity to treat the condition successfully. In other words, it is important that the screening process leads to demonstrable health gains from the earlier detection afforded by screening. Second, in addition to detecting true positive cases, a screening technology may erroneously identify some individuals as positive who do not have the underlying condition or fail to identify people who genuinely have the disease. All of these false positives and false negatives may incur additional costs and impose significant negative health effects.

Life-years gained
In circumstances where screening may prevent premature mortality, another useful measure of benefit can be expressed in terms of the number of additional life-years gained from the application of screening. The focus should be on the incremental life-years gained; that is, the difference between the future life-years expected for those detected presymptomatically minus the future life-years expected for those detected symptomatically who survive.

For screening programmes, these calculations should net out any mortality effects that may result from a significant volume of false-negative cases, and take account of any generally lower life expectancy for those affected by serious disabilities.

A major limitation of using incremental life-years gained as a measure of benefit is that no account is taken of the quality of the future life-years gained. In many cases there may be systematic differences between the quality of life of presymptomatic and symptomatic cohorts. The latter may be susceptible to poorer health generally, and a proportion may have already suffered physical and neurological insults with permanent consequences for HRQoL.

The assessment of health benefits using life-years gained requires that sufficient information is available about the natural course of the condition, which will allow the identification of how many additional life-years may be gained through early detection and treatment. For some of the more recently identified metabolic disorders, evidence on future life expectancy with individual conditions remains limited.

QALY gained
Given the potential impact of inherited metabolic disorders on the future HRQoL of affected individuals, and the evidence that there would be systematic differences between screened and unscreened cohorts, the use of a QALY framework would provide the most robust basis on which to quantify the health benefits of neonatal screening. Unfortunately, there is currently no published evidence on what might be appropriate health-status weights to use within a QALY framework for the various metabolic disorders considered here.

There is evidence that for some of the conditions considered here (MCAD deficiency, in particular), overall quality of life is near normal for the vast majority of cases if treatment is initiated early enough.[63] There may be future acute problems related to episodes of decompensation after diagnosis is made, but these postdiagnosis episodes would not vary systematically according to presymptomatic or symptomatic detection.[60] This suggests that the observed incremental difference in calculated QALYs for MCAD deficiency would depend to a large extent on the proportion of symptomatic patients who develop permanent disabilities and impairments.

For other conditions, however, the difference between early detection/treatment and subsequent

morbidity is less clearly defined. In some cases, for example, developmental problems may be present regardless of the age of diagnosis or treatment onset.[84] Morbidity may also remain high despite treatment. A recent review of 50 cases of long-chain 3-hydroxyacyl-CoA dehydrogenase deficiency concluded that "morbidity remains alarmingly high despite current therapeutic regimes".[85] The absence of HRQoL data for some of the conditions, for both treated and untreated cases, represents a significant limitation on the overall evaluation of screening benefits.

Other non-resource benefits

Other non-resource benefits attributable to neonatal screening would be effects on future quality of life for parents and families caring for children and young adults affected by serious disabilities or other cognitive impairments. Note that no assumption is made about whether these effects overall would be positive or negative. Many parents of children with disabilities identify major benefits from the care relationship they develop with their children and consider that, on balance, the quality of their life has been enhanced, not diminished.

A further non-resource effect may be the avoidance of anxiety and distress caused to families with children who have acute episodes of illness before symptomatic diagnosis is made; this may be accompanied by additional anxiety and distress caused by the implicit misdiagnosis involved in repeated symptomatic episodes of acute illness. Loss of a child as a result of a preventable metabolic crisis will also have profound effect on the quality of life and welfare of affected parents and families.

Some of the non-resource effects also specifically relate to the processes and outcomes of the actual screening programme. These arise in a number of ways.

As a consequence of testing

These effects derive from both the initial screening intervention and any subsequent confirmation tests performed.

As a consequence of diagnosis

Effects here are related to the receipts of a correct diagnosis and also occur as a repercussion of diagnostic errors produced by the screening technology. The anxiety and distress caused by a correct diagnosis will occur either at the presymptomatic stage (through screening) or when the infant is diagnosed symptomatically at a later time.

Diagnostic errors occur as a result of either false-positive or false-negative diagnoses that arise as part of the screening process. The overall size of these negative effects due to misdiagnosis tends to be directly related to the sensitivities and specificities of the screening technology used.

Given the difficulties and complexities of diagnosis in the absence of a robust screening technology, diagnostic errors also occur (in some cases repeatedly) during symptomatic presentation of affected individuals.

It should not be assumed that all of the above effects are negative. For example, parents often derive positive benefits from the receipt of a definitive diagnosis, particularly in circumstances where they feel that they have had difficulties in having their concerns taken seriously or have not been able to explain adequately their child's recurrent illness. Related to this there is also evidence that a proportion of infants' deaths previously classified as SIDS may now be explained definitively by an inborn metabolic disorder.[86,87]

All of the above non-resource effects are likely to be significant. To date, however, no published studies have attempted to quantify these non-resource consequences and externalities, using a health utilities framework, which would allow direct comparison of positive and negative effects, in respect of neonatal screening for inborn errors of metabolism.

However, several studies have examined the psychological impact of both neonatal and other screening programmes, which is of relevance to an assessment of some of these non-resource effects identified. A systematic review of the available literature was performed as part of one of the previous HTA reviews.[1] This review examined the evidence on two related aspects of relevance to neonatal screening: first, the direct psychological impact of screening on both infants and their parents; and second, how any impact of diagnosis provided through screening could be managed to reduce possible anxiety and distress.

The main conclusions of this review of the psychological impact of screening are worth repeating here:

> "The systematic review of the research surrounding the psychological impact of neonatal screening indicates that the psychological benefits of neonatal screening outweigh the costs. This can be seen to be the case for the psychological effects of the medical procedures involved in screening, as well as for those

involved in the early diagnosis of genetic disease through screening in comparison to traditional clinical diagnosis."[1]

But the review also concluded:

> "There is evidence to suggest that parental anxiety and distress following diagnosis through screening relates to the type and amount of information provided to parents as well as to the actual methods by which information is conveyed to them."[1]

The latter clearly has some implications for the organisation and management of both the current neonatal screening programme and any future extension of neonatal screening to include other metabolic disorders.

Note, however, that it was not an objective of the present review to update or re-examine recent evidence on the psychological impact of neonatal screening on infants and their families.

Review of literature: methods

Search strategies

Sources searched

In addition to the 12 electronic bibliographic databases covering biomedical, science, economic and grey literature for tandem MS and individual inborn errors of metabolism, separate searches in NHS Economic Evaluation Database (EED) and the Office of Health Economics (OHE) Health Economic Evaluation Database (HEED) were conducted to identify other published studies that would have economic relevance to tandem MS or the neonatal screening of inborn errors of metabolism (Appendix 37). The reference lists of relevant articles and abstracts of conference proceedings were checked and various health services research-related resources were also consulted via the Internet. The main focus was on identifying papers that included reference to economics, costs, outcomes and quality of life.

Search terms

Combinations of free text terms were used. Further details are provided in Appendix 37. In addition, methodological search filters designed to retrieve economic evaluations and quality of life studies were applied to the searches on tandem MS and the major inborn errors of metabolism in MEDLINE (Appendix 38).

Search restrictions

Searches were restricted to publications after 1995 to the present (January 2002), as previous economic evaluations would have been covered in the existing HTA reviews.[1]

Inclusion and exclusion criteria

The titles, and abstracts where available, of all the articles identified by the literature searches were downloaded into a database. Any duplicates were removed. Papers were subsequently excluded for the following reasons:

- foreign language paper
- case reports based on fewer than two patients.

Although there is general guidance on the methodological quality of published studies of relevance to the review of the economic literature, more specific guidance on assessing the quality of economic evidence is available. In addition to specific recommendations, some provide checklists to assist with the methodological review of economic papers.[88,89]

Review of the literature: results

Economic evaluations of neonatal screening for inborn errors of metabolism

Systematic reviews of published studies before 1995 on economic evidence appropriate to an assessment of neonatal screening for inborn metabolic disorders were contained in two previous reports commissioned as HTA reviews.[1,2]

Pollit and colleagues[1] undertook cost-effectiveness modelling of a possible extension of neonatal screening to other inborn errors of metabolism using tandem MS. The implicit viewpoint of the modelling exercise was that of the NHS, although costs (at 1996 prices) were restricted to those incurred as part of a laboratory screening process and the (net) impact on treatment costs related to the screening or no screening options considered. Benefits were measured in terms of number of cases detected and estimated additional life-years gained as a consequence of earlier detection and treatment of inborn errors of metabolism. The results were expressed in terms of incremental cost per true case detected and incremental cost per life-year gained. Single parameter sensitivity analysis was performed on key variables such as tandem MS costs, treatment period and avoided mortality.

The report concluded that the decision analysis tentatively supported the introduction of tandem MS, provided that additional disorders were

included as part of the neonatal screening programme (in accordance with the priority ordering suggested by the economic modelling). However, it recommended that a pilot introduction of tandem MS should be undertaken, supported by a prospective gathering of evidence, sufficient to enable a full economic evaluation of neonatal screening using tandem MS.

The Seymour report[2] also provided a systematic review of the published literature. No formal economic modelling or primary research was undertaken as part of this review. The report concluded that evidence on the use of tandem MS was limited, and at best a case could only be made for the use of tandem MS for screening for MCAD deficiency and GAI. It recommended that a prospective evaluation be undertaken including the estimation of screening costs and cost-effectiveness, before tandem MS technology and an extension of neonatal screening could be adopted as national policy. Since the two previous HTA reviews, no studies have been identified which seek to provide an economic evaluation of the costs and benefits of screening for inborn metabolic disorders using tandem MS. (The authors were aware that a number of economic evaluations on the use of tandem MS for screening of inborn errors of metabolism were underway, but these studies were not published in time to be included within this review.)

Only one formal economic evaluation of neonatal screening in the UK has been published since the HTA reports of 1997. This evaluation[90] was based on an assessment of the cost-effectiveness of screening for PKU using a compilation of secondary economic data in a decision model. Costs included were sample collection, laboratory testing and repeat tests, the costs per case of dietary therapy and other follow-up care costs. Inclusion of estimated lifetime treatment costs was important, and explicitly addressed a comment in one of the previous HTA reviews that this omission was a significant deficiency within existing knowledge about the cost-effectiveness of PKU screening.[2] Benefits were expressed in terms of future resource effects rather than any direct health effects that might be obtained (e.g. life-years and HRQoL). Benefits, therefore, comprised avoided future health and social care costs related to disability caused by the condition and "avoided loss of productivity for PKU patients and their parents".[90] In accordance with recommended good practice, the calculated indirect costs were quantified separately. Although the quality of the secondary data sources used was variable (and explicitly recognised by the authors), extensive sensitivity analysis was performed. This comprised the use of median, best and worst case scenarios. In addition, a best estimate based on probabilistic sensitivity analysis using a Monte Carlo simulation technique was calculated. Assumptions used within the probabilistic model were stated and prior distributions defined and presented. The study concluded that:

> "Screening for PKU is cost saving in the UK, and on its own justifies the cost of collecting neonatal blood samples, which are also used for other screening tests. This result is robust to changes in assumptions about treatment and outcomes".[90]

Although of value to the current review, this study was confined to estimating whether the existing screening programme for PKU alone could be justified. No attempt was made to assess the relative cost-effectiveness of using tandem MS for PKU screening or additional inborn errors of metabolism.

This is the only recent publication to date (prior to January 2002) that has specifically sought to provide an explicit economic evaluation of a neonatal screening programme for inborn errors of metabolism. Therefore, systematic reviews of the economic evidence on the use of MS/MS for the UK neonatal screening programme have provided no tangible additional evidence on the potential use of tandem MS.

Several papers have been published that refer to cost, treatment and outcomes. Many of these, however, simply reference previous research or reiterate conventional understanding on these issues. Some provide new data, but do not give explanations of the basis on which this type of data was derived that allow appropriate assessment. The quality and potential generalisability of the results obtained are too uncertain to be of value. As an example, one study of very long-chain acyl-dehydrogenase deficiency indicated that presymptomatic diagnosis would have prevented "at least part of a lengthy and intensive prediagnosis hospitalisation that had an estimated cost of $400,000".[91] However, in addition to basing the study on a case report of only one individual (itself a reason for exclusion), no information on how the costs were estimated is given.

Many of these secondary papers, therefore, do not attempt a systematic economic consideration of the costs or benefits associated with either presymptomatic or symptomatic detection of inborn metabolic disorders and are therefore excluded from the review.

In conclusion, no economic evaluations have yet been published which provide detailed information on the economic benefits and costs of replacing existing technologies for PKU screening with tandem MS, or of extending neonatal screening programmes using tandem MS to include additional conditions such as other amino acid, acylcarnitine and urea cycle disorders.

Additional published evidence of relevance to the economics of neonatal screening

Although no formal economic studies have been published in recent years, several other recent studies supply information and results that are of relevance to an overall assessment of the economic consequences of screening for inborn errors of metabolism using tandem MS. Although these studies are not systematic economic analyses, they indicate the potential health effects and resource consequences of either presymptomatic or symptomatic detection of additional neonatal metabolic disorders. In the context of the present review, these publications provide information that assists with the construction of the parameter values used in the economic model described later in this report.

A good deal of this research has focused on MCAD deficiency as a condition with "relatively high incidence, significant levels of mortality and morbidity if undetected early and effective and uncomplicated treatment strategies".[21]

Two specific aspects of this literature are identified which provide significant data of relevance to an economic modelling of screening for inborn errors of metabolism: first, the extent to which symptomatic presentation of the disorders incurs additional acute healthcare costs which may be avoided through presymptomatic detection, and second, the extent to which, for some of the conditions considered here, symptomatic presentation may be attended by physical and neurological insults, which may lead to significant disability and impairment in affected individuals. Again, presymptomatic diagnosis and early treatment may avoid the additional future healthcare and social care costs associated with these disabilities.

Evidence on MCAD deficiency indicates that symptomatic diagnosis may be attended by a number of acute hospitalisation episodes before correct diagnosis is made. In one study of 36 patients, 17 had been admitted to hospital with symptoms characteristic of MCAD deficiency before correct diagnosis. Of these patients, eight had previous admissions with coma (three more than once) and a further nine had previous admissions for lethargy and/or hypoglycaemia (six of these more than once).[63]

Some information on the possible disabilities associated with symptomatic presentation of MCAD deficiency has also been published. In addition to an earlier US study on outcomes in a cohort of 120 cases,[60] a more recent UK study indicates the extent to which impairments may arise through symptomatic presentation.[63] In the latter study, of 30 remaining symptomatic patients (3 months after discharge from hospital), one patient had mild cognitive impairment, two patients had moderate impairment and another two had severe impairment. One group of paediatricians in the UK has indicated that between 10 and 15% of patients with MCAD deficiency are left "severely handicapped".[92]

Other studies have indicated the extent to which symptomatic presentation of GAI may result in significant future impairments.[68,69] In addition, there have been several publications of relevance to the economic assessment of screening for homocystinuria (due to CβS deficiency). However, published evidence on the natural course and long-term outcomes for some of the rarer metabolic conditions remains limited or absent.

Economic modelling

Introduction

It is clear from the review presented above that there has been very little additional research published since the previous HTA reviews that provides any robust economic evidence on the relative cost-effectiveness of neonatal screening for inborn errors of metabolism, and no studies specifically relate to the use of tandem MS. A number of cost-effectiveness studies (outside the UK) are currently underway, but the results are not yet available. In this context, therefore, it was decided to undertake an economic modelling exercise that would provide some update of the work previously undertaken in one of the earlier HTA reports.[1]

The primary purpose of this modelling was to indicate the likely effects of introducing tandem MS in the current UK screening programme for PKU and how the ability of this new technology simultaneously to screen for other conditions might affect the decision.

This was felt appropriate for two reasons: the costs of using tandem MS technology may have changed significantly in the intervening period; and some additional research information is now available on the application of tandem MS for detecting inborn errors of metabolism (e.g. sensitivity and specificity for particular disorders). Further research has also been published on the diagnosis, treatment and consequences of various inborn metabolic disorders, which may be of relevance to the economic modelling.

The economic modelling undertaken here, however, was not simply designed to replicate the work presented previously,[1] adapted to reflect updated costs. An important aspect of the work described here is an attempt to capture where uncertainty around information exists and then to seek to quantify the consequences of this uncertainty on the decision about the use of tandem MS. This has the benefit of indicating where any future research, if required, should concentrate to reduce the uncertainty around those parameters that would have the most significant effect on the relative cost-effectiveness of using the newer technology.

The approach presented here adopts a Bayesian probabilistic framework applying Monte Carlo simulation procedures within an EVPI modelling tool developed by the Operational Research Department, School of Health and Related Research, University of Sheffield. Probabilistic analysis was undertaken by assigning prior distributions to represent the uncertainty that surrounds model inputs.

All parameters applied, and the distributional assumptions used to characterise the uncertainty around each parameter, are set out in Appendices 34–36 and 39–44. Where appropriate, additional justification for parameter values and distributions used is provided in the following sections.

In addition, Appendix 45 provides a summary of the parameter inputs used for both the PKU and MCAD deficiency models described later. These not only identify all the models inputs used, but also show formulae to indicate the relationship between elements of the model and how the cost and benefit outputs for each simulation were calculated. The actual parameter values shown are illustrative only; these values change during simulations according to the prior distribution used.

Appendix 45 (*Table 18*) provides information about the phase i comparison between existing and tandem MS technologies for PKU screening. Appendix 45 (*Table 19*) shows model inputs and relationships for the (PKU+) MCAD deficiency model.

Methods

The economic viewpoint adopted was explicitly that of the health and other public sector providers within the UK. Costs comprised screening and treatment costs associated with a particular condition. Estimates were also included of acute care costs related to symptomatic presentation and future healthcare, social care and education sector costs of moderate to severe disabilities caused by some of the disorders. Patient-related and informal carer costs were omitted, as were estimates of any (net) future changes in productivity. In the absence of any estimates of quality of life weights for the metabolic disorders considered, benefits were expressed in terms of life-years gained. The economic modelling performed was undertaken in three distinct phases.

The first phase (i) evaluated the consequences of replacing the existing PKU programme with the newer technology, tandem MS. This involved comparing the estimated costs of screening for a hypothetical cohort of 100,000 neonates using a combination of the existing technologies (Guthrie, fluorometry and chromatography) used in the current screening programme, with the estimated cost of undertaking the same work using tandem MS. Existing studies already indicate the cost-effectiveness of screening for PKU in the UK;[90] therefore, the study comparators for phase i are valid.

Essentially, as the health benefits of detecting PKU through neonatal screening were assumed to be identical irrespective of the technology used, this element of the economic modelling largely identified what would be the incremental costs incurred or saved from the change to tandem MS for PKU. More information about the model assumptions used is given below.

The second phase (ii) of the modelling was designed to indicate a priority ordering in the possible extension of neonatal screening for additional conditions. By sequentially testing the impact of adding a screen for a particular disorder to the use of tandem MS for the PKU programme, it was possible to derive a value for the cost per additional life-year gained for each condition. This is based on combining both the estimated incremental cost/saving of using tandem MS for

PKU with any incremental costs for presymptomatic identification of the individual disorder, and dividing this overall additional cost by the estimated additional life-years gained.

The final third phase (iii) in the modelling is based on the overall effects of adding extra disorders screened to a PKU programme based on tandem MS. For example, if the application of tandem MS to PKU screening alone were not cost-effective, what effect would the addition of the next priority disorder (estimated in phase ii) have on the cost-effectiveness of neonatal screening using tandem MS? This is done by combining any incremental costs/savings attributable to using tandem MS for PKU with the additional costs (net) and benefits (life-years gained) of adding the new disorder to the screening programme.

Economic modelling was undertaken using Monte Carlo simulation methods with parameter distributions defined according to the evidence presented in this and previous systematic reviews of inborn metabolic disorders. In addition to providing conventional data on the ICER obtained from a probabilistic analysis of model inputs, this framework is used to estimate the expected costs (opportunity losses) associated with parameter uncertainty.

Modelling assumptions

PKU screening: existing technologies and tandem MS

Information on the estimated incidence of PKU for the UK was based on information obtained from the previous and current literature review. The base-case estimated incidence for all PKU cases in the UK was 9 per 100,000 neonates. This was introduced into the model as a Poisson distribution (a statistical distribution relevant to the modelling of relatively rare events).

A significant element in the total costs of providing a neonatal screening service relates to the costs of obtaining and transporting samples to the laboratories. The 1997 HTA report[1] estimated the cost of collecting a first sample at £5.04 (1996 prices), while another review obtained a broadly similar estimate of £4.82 based on 1995 prices.[90] The current cost of an initial sample was estimated using the same staff and consumables cost estimated within the 1997 HTA report and revalorised to 2001 prices using the NHS Health Services Cost Index. Thus, the cost of collecting an initial blood sample for screening was calculated to be £5.65 per specimen.

In addition to the cost of collecting and dispatching the first samples it was expected that about 1% of all initial specimens would be found to be inadequate, for a variety of technical reasons unrelated to the screening technology used in the laboratories.[1] These extra specimens would require dedicated visits by midwives or health visitors. The cost of obtaining a repeat sample was estimated as equivalent to the cost of a dedicated home visit by a midwife or health visitor, plus an allowance for consumables. The current cost of a single visit by a midwife (£16) and a health visitor (£25) was obtained from 'Unit costs of health and social care: 2001,[93] published by the Personal Social Services Research Unit at the University of Kent. This is an annual publication that provides up-to-date information on nationally applicable costs and is used extensively in the costing of healthcare packages and labour inputs from health professionals in the NHS. From this, the estimated cost for an additional collection of a specimen was £20.15 (assuming that midwives collect two-thirds of all repeat specimens and that consumables cost £1.15).

However, as these costs are identical whether the PKU screening programme uses existing methods or tandem MS, they can be excluded from the evaluation of the cost-effectiveness of adopting the newer technology for PKU screening alone. Note that this situation will not necessarily occur if neonatal screening is extended beyond the existing programme.

Nevertheless, the basic specimen collection process (i.e. not related to confirmation protocols) incurs significant real resource costs that should not be overlooked. Specimen collection costs are therefore estimated to be about £565,000 for each cohort of 100,000 neonates screened. This estimate excludes the cost of collecting additional samples due to the technical insufficiency of the initial specimen; these costs are included within the evaluation of screening technologies because there will be systematic variations related to differences in the unit costs of laboratory screening per specimen.

It was not the intention of this review to commission any new primary research. Although contact was made with appropriate suppliers of tandem MS technology and with the Sheffield Neonatal Screening Laboratory, information necessary for the economic modelling relied mainly on secondary data. A systematic approach to all existing screening laboratories to establish their current costs of neonatal screening for PKU was not undertaken.

TABLE 10 *Estimated cost of existing screening technologies for PKU*

Method	Original data[a]			Revalorised data	
	No. of laboratories	Cost per neonate (£.p)		Cost per neonate (£.p)	
		Median	Range	Median	Range
Guthrie	6	0.64	0.55–1.45	0.74	0.63–1.67
Fluorometry	4	1.08	0.51–1.36	1.24	0.59–1.56
Chromatography	5	0.42	0.31–0.93	0.48	0.36–1.07
		Probabalistic estimate: mean = 0.92			
		Cost per neonate for all technologies combined:[b] N(0.92, 0.12)			

[a] Based on Pollitt *et al*. (1997).[1]
[b] Median derived from probabalistic estimate: 0.92.

Estimated costs for the use of existing technologies for PKU screening were based on the summary information presented in the previous HTA review.[1] The summary was compiled from a detailed cost questionnaire sent out to all laboratories at that time. As the current neonatal screening programme is based on the application of three different methods (Guthrie, fluorometry and chromatography), it was decided to derive a probabilistic estimate of the cost per sample screened for all three screening technologies combined. This was based on the sample costs derived in 1996 and the number of laboratories using each technology identified in the cost questionnaire returns.

The range of cost elements identified and measured is consistent with the resource elements used to cost the tandem MS screening technology in this report. However, it was not possible to assess in detail how data obtained from individual questionnaires from 1996 were costed. It is clear that there were wide variations in reported costs obtained from these schedules, even for similar screening methods. These variations will partly comprise genuine operational differences in efficiency between laboratories (different combinations of inputs, scale, etc.), and some may partly reflect differences in the reporting of costs.

A smaller simulation exercise was carried out to capture the full range of costs reported for existing technologies in the previous HTA report. This involved describing a triangular distribution of cost for each method based on the median (as 'most probable') cost per specimen reported. All cost per specimen values reported in the HTA document were revalorised to current prices using the Health Services Cost Index. An estimated mean value was then calculated based on a weighted average of the results from each distribution (i.e. weighted by the number of laboratories using a particular technology). The simulation results obtained (using 10,000 iterations) are presented in *Table 10*.

The resulting mean (and standard deviation) values from this exercise, presented in *Table 10*, were then included in the phase i model using a normal distribution as a best estimate of the cost per sample using all existing technologies.

Note that using estimates based on the distribution of costs from the previous HTA report implicitly assumes that the costs associated with the existing technologies option have the same relative combination of screening modalities and operating efficiencies as obtained in 1996.

Information on tandem MS systems and ancillary equipment and consumable costs were obtained from two major suppliers of laboratory equipment in the UK.

Estimates of annual equipment maintenance costs were also obtained from suppliers. Overhead facility charges for the accommodation of a tandem MS screening method were estimated from the data originally obtained as part of the cost questionnaire returns obtained for the previous HTA report.[1] These comprised a range of estimates of the laboratory maintenance costs-based floor area devoted to neonatal screening. There were several reasons for using these data. First, as tandem MS does not require any special environmental conditions it was assumed that a tandem MS method in total (including storage and preparations areas, etc.) would occupy a similar amount of accommodation in a laboratory as existing technologies. For consistency with

existing technology costs, it was felt appropriate to obtain estimates based on the full area related to neonatal screening, not just the limited space occupied by a tandem MS machine. Second, using information on accommodation used at a single site may be significantly unrepresentative. An annual allowance for facility costs was therefore obtained using the distribution of estimates previously attributed to neonatal screening for existing technologies, again revalorised to current prices using the Health Services Costs Index.

Labour inputs required for a neonatal screening laboratory that uses a tandem MS system were estimated in consultation with the Neonatal Screening Laboratory at Sheffield Children's Hospital (which currently uses tandem MS). Labour inputs for a single system were estimated to be 1.0 whole-time equivalent (WTE) of a Medical Laboratory Scientific Officer (grade 1) for management of the specimens collected and analysed (e.g. preparation and loading of microtitre plates). An estimated additional 0.2 WTEs of a Clinical Scientist grade would be required for reading and analysis of screening results. Both labour inputs were costed using the midpoint of appropriate NHS salary scales with an additional allowance of 15% for employers' on-costs (e.g. National Insurance contributions). These costs are set out in Appendix 46.

Recognising that a number of important parameters can affect the cost of a tandem MS system per sample screened, it was decided to conduct another simple probabilistic exercise to estimate unit costs. Three parameters in particular have a significant impact on cost per specimen. First, there is the magnitude of the initial capital outlay itself. The current capital cost for a single tandem MS system suitable for screening inborn errors of metabolism ranged from about £165,000 to about £180,000. Including all ancillary equipment (e.g. nitrogen concentrator) the economic model assumed a most probable capital outlay of £175,000 for a complete tandem MS system, but varying between the limits indicated above (see Appendix 47).

This initial capital outlay for tandem MS was converted to an annual equivalent cost, based on the expected life of the asset and using a discount rate of 6%, in accordance with accepted practice for the treatment of major capital purchases in economic evaluations.[77] The discount rate was not varied.

A second important consideration is the expected life of the asset. The current NHS Capital Accounts Manual advises that capital equipment should be depreciated over its "useful economic life". This is a concept that now allows some discretion over what constitutes the useful economic life of an asset. However, the manual suggests that costing procedures should still adhere to the standard lives of equipment assets as laid down in earlier guidance unless significant adjustments are warranted. The standard life for medium-term medical equipment is set at 10 years. The manufacturers of tandem MS, however, felt that a much shorter period (5 years) was more realistic. For the purposes of the economic model, an asset life of 7 years was assumed as a most probable intermediate position.

Third, and most significantly, unit costs depend on the volume of samples processed in the course of a single year on one machine. With capital-intensive technologies, the higher the volume of output processed using the equipment, the lower the average unit costs obtained because the fixed (capital) costs are distributed over more units of output. Although many tandem MS systems in use are capable of handling between 500 and 800 samples per day, allowance needs to be made for inevitable redundancy (e.g. downtime for maintenance, preparation and cleaning). In addition, individual laboratories participating in a neonatal screening programme are unlikely to achieve throughput rates anywhere near the maximum capacity.

As the number of specimens processed on a single system not only affects the cost per sample obtained but also has important service delivery implications for the number of laboratories participating in a neonatal screening programme, it was decided to estimate a cost per specimen for a range of different processing volumes. The volume levels chosen were 20,000, 40,000, 50,000, 60,000, 70,000 and 80,000. In addition, a probabilistic estimate based on an operational range between 50,000 and 60,000 samples per year on a single instrument was calculated.

Although a single tandem MS system has the technical capacity to process more than 80,000 samples in a year there are operational reasons why volumes above this level are not considered in the modelling exercise. It is likely that a neonatal screening programme that expected average processing volumes per system above 80,000 would involve a significant step increase in costs. In part, this would be due to the extra staffing needed to manage such a large volume of samples per year in a single laboratory. In addition, at

TABLE 11 *Cost per sample for different processing volumes of tandem MS (probabilistic estimate of cost per sample using tandem MS technology)*

Volume of specimens per tandem MS system	Cost (£.p) per specimen	
	Mean	Distribution
20,000	3.96	N(3.96, 0.225)
40,000	2.00	N(2.00, 0.113)
50,000	1.61	N(1.61, 0.089)
60,000	1.35	N(1.35, 0.076)
70,000	1.17	N(1.17, 0.064)
80,000	1.03	N(1.05, 0.056)
Range estimate: 50,000–60,000 samples	1.45	N(1.45, 0.10)

volumes above 80,000 per annum on a single system, it would be necessary to consider having a second instrument for back-up; without this it may be extremely difficult for a laboratory to catch up after stoppages.

Data in *Tables 11* and *12* show the costs associated with using a much lower volume level (20,000 samples per system). However, it would be inefficient to use such a capital-intensive method at operating volumes so far below capacity. The implications for service delivery and the organisation of laboratory screening services are considered as part of the discussion of the model results. Therefore, 40,000 samples per year was regarded as a realistic lower limit, particularly when the previous HTA report concluded that, because "laboratories undertaking tandem MS screening should have adequate workload to justify the capital investment and staffing required", a minimum of 50,000 samples per year was suggested.[1]

For the purposes of assessing changes in unit costs per specimen analysed over the range of volumes considered here, labour and some of the other annual expenses (overheads, maintenance and internals standards) are treated as fixed costs (and characterised using simple normal distributions). Reagents are input as a direct variable cost and estimated to vary between £0.03 and £0.05 per sample.

All capital, consumable and labour cost values used and the distributions applied for the derivation of the probabilistic estimates for each of the operating volumes identified are presented in Appendix 47. The estimated cost per specimen results obtained using a tandem MS system for different workload levels are presented in *Table 11*. These values, and the associated distributions shown, were then used in the economic model to evaluate the impact of using tandem MS for PKU screening compared with the combination of existing technologies.

Confirmation costs for all positive PKU cases identified by screening (including false positives) were based on estimating the laboratory costs related to the confirmation protocol followed for PKU. This combined information on the additional laboratory tests required as part of the protocol and information on charges for these tests made by other laboratories supplying services to a neonatal screening laboratory. This was based on information derived from the Trent Neonatal Metabolic Screening Laboratory Service (Pollitt R: personal communication; 2002).

It was assumed that any repeat sampling needed as part of the confirmation protocol would require an additional contact with the neonate. The cost of this additional contact (£19) is based on a combination of dedicated midwife and health visitor visits (two-thirds and one-third, respectively). These costs were valued using the annual publication 'Unit costs of health and social care: 2001',[93] plus a small allowance for consumables (£1.15).

All laboratory tests, and the cost of repeat specimens, were combined into a single confirmation protocol cost (£60) applied within the model (Appendix 39). The distribution of these costs was characterised using a log-normal distribution.

The cost of referral and advice, identified separately in the previous HTA report but included in the overall cost per sample values for existing

technologies, was included as a fixed additional staff cost within the estimated cost per sample using tandem MS (see Appendices 45 and 46).

Repeat sampling due to technical problems with the samples provided was assumed to be independent of the screening technology used. Therefore, the estimate of 1% repeat sampling for technical reasons[1] was applied to both existing and tandem MS technologies.

The sensitivity and specificity (and therefore the false-negative and false-positive rates) of both existing and tandem MS technologies were estimated from the literature and the original HTA model.[1] Information on the number of false negatives produced by both technologies was limited. The model adopted a conservative assumption that the false-negative rates were the same for existing and tandem MS screening modalities (0.02% of screened cohort).

As the overall false-negative rate was assumed to be the same for both options, all future treatment costs and morbidity outcomes were assumed to be the same for both the existing and new technology. Therefore, all further costs were excluded from the phase i model. All parameters and distribution assumptions used in this phase i PKU model are set out in Appendix 39.

Priority ordering for additional conditions

To obtain a priority ordering of which conditions should be considered first for addition to the neonatal screening programme, the economic modelling undertaken within the previous HTA report was reworked. However, two important factors have made this work different. First, treatment and confirmation costs relevant to each disorder were updated to 2001 prices. Second, and more importantly, the incremental costs associated with using tandem MS for PKU (identified in phase i) were included in the assessment of each condition separately to establish which would have the largest initial impact on (net) changes to costs and health benefits.

In addition to the incremental cost of using tandem MS for PKU, an important driver of cost associated with screening for each disorder is related to differences in treatment as a consequence of presymptomatic detection. As indicated previously, information on the treatment strategies and costs for each condition was taken from the 1997 HTA report;[1] with prices revalorised to 2001 using the Health Services Cost Index. Details of the specific treatment profiles adopted for each condition were also the same as those originally defined in the 1997 HTA report.[1] This information is reproduced in Appendices 34–36. Treatment regimens for particular disorders are categorised according to five broad groupings and used to estimate the future costs related to each condition. The revalorised costs for each category are presented in Appendix 35.

It is acknowledged that these treatment strategies may not now capture completely what represents current practice for some conditions. Given the uncertainty surrounding the validity of these treatment regimens and the costs associated with each, the model included a variable that allowed future treatment costs to vary from base-case estimate (as set out in Appendices 34–36) and twice base-case estimate. The exception was tyrosinaemia, where only the dietary cost elements were varied, not the cost of a liver transplantation.

For those conditions not expected to receive lifetime therapy (MCAD deficiency, branched-chain and long-chain defects), costs were entered into the model using a uniform distribution varying from the base-case estimate to four times the base-case estimate. This was designed to characterise variations in practice over the number of years that treatment should be prescribed for these specific disorders and uncertainty about the actual resource cost of the profiles used. A three-fold magnitude of variation was employed in the previous HTA report for these specific conditions as part of the sensitivity analysis for these conditions,[1] but not introduced in a probabilistic fashion as here.

These models also assumed at least one acute hospitalisation for each symptomatic presentation as an avoidable cost of screening. In addition, future healthcare and social care costs related to disability were included where there was evidence to indicate what proportion of symptomatic cases could be affected. It was felt appropriate to include these estimates if available so that the priority ranking would correctly identify which condition should have the greatest immediate impact on the use of tandem MS for PKU screening.

For the purposes of obtaining a priority ordering of the additional conditions, most of the data on incidence, proportions of cases affected, the diagnostic accuracy of tandem MS and the assumptions on mortality effects of presymptomatic detection were taken from the 1997 HTA report.[1] Where new information was available from the

literature review undertaken as part of this project (e.g. data on outcomes following treatment, sensitivity and specificity of tandem MS and effectiveness of screening), this information was used. In addition to MCAD deficiency, specific new information of relevance to the economic modelling was available for GAI and homocystinuria due to CβS deficiency and this was included in models for these conditions (see Appendices 41 and 42). The treatment category, incidence and health-effect assumptions used for all other inborn metabolic conditions in phase ii of the modelling are presented in Appendices 36, 43 and 44.

Extending PKU screening to additional conditions using tandem MS

The results of the previous analysis indicated that the highest priority should be accorded to an extension of screening for MCAD deficiency.

MCAD deficiency

To evaluate the effects of adding MCAD deficiency screening to a PKU programme run using tandem MS technology, additional modelling was undertaken using Monte Carlo simulation. The following base values and distributions were defined and included within the model.

The true birth prevalence/incidence of MCAD deficiency in the UK varied significantly within the literature reviewed: from approximately 4 to 10 per 100,000. It was therefore decided to retain the estimate used in the previous HTA model. Thus, an overall expected incidence for the UK for MCAD deficiency of about 8 per 100,000 was used. It is recognised that this estimate of cases detected through screening will include a proportion that would remain asymptomatic in the absence of screening. As cases detected are relatively rare events, incidence per cohort of 100,000 neonates was characterised in the model using a Poisson distribution.

Information on the sensitivity and specificity of screening was obtained from the literature reviewed. The false-positive rate was estimated to be 0.0230% (95% CI 0.0159 to 0.0297%). The evidence suggests that neonatal screening using tandem MS has produced no false negatives so far. However, this was regarded as potentially overoptimistic and therefore it is assumed that tandem MS will generate one false negative for every one million samples screened.

Treatment profiles and costs, for those detected presymptomatically and those detected symptomatically, were based on those prescribed for MCAD deficiency as part of the previous HTA economic model.[1] As indicated earlier, however, because of uncertainties around estimated MCAD deficiency treatment costs based on the previous HTA report strategies, the costs applied to screened and unscreened cohorts were characterised as a continuous uniform distribution between the estimated base-case value and four times the base-case value.

It is assumed that all cases of MCAD deficiency detected by screening would receive treatment, whereas only the proportion that would become symptomatic[20] would receive treatment in the absence of screening.

Three aspects of the current modelling provide useful extensions to that previously undertaken as part of the HTA 1997 review.[1] First, although screening for additional disorders using tandem MS can be undertaken using the same initial blood specimen obtained as part of the PKU programme, it was not assumed that the marginal cost of screening for an additional disorder was thereby zero. The following costs were added and included.

- An estimate of the extra time taken by a midwife or health visitor in explaining the implications of screening for additional disorders and obtaining consent was added. This also includes the costs of providing amended/further explanatory materials for parents. This cost was estimated (in consultation with a midwife) to add about £0.30 per specimen collected.
- For a relatively frequent condition such as MCAD deficiency, approximately 0.1 WTE of a clinical scientist grade is required to deal with additional reporting, referral and advice. This was estimated at about £3500 as an extra fixed cost for a single neonatal screening laboratory.
- A further £0.10 per sample was added to account for additional reagent costs, internal standards and other consumables.

Second, the model developed here includes some estimate of the acute healthcare costs associated with symptomatic detection of MCAD deficiency. The previous HTA report only included the incremental treatment costs related to earlier detection through screening, but did not assess the healthcare costs imposed by those episodes that led to symptomatic diagnosis.

The base-case assumption made within the current modelling exercise was that symptomatic

presentation would incur, on average, at least one acute hospitalisation episode that precipitates diagnosis in the absence of a screening technology although, as previously acknowledged, there may be some occasions when there are no symptomatic acute episodes. For MCAD deficiency, there is evidence that many patients have also experienced previous acute hospitalisations before diagnosis.[63] Therefore, a base case of one acute symptomatic episode was included in the model, but with the distribution characterised using a Poisson distribution. For MCAD deficiency this is taken to be a conservative assumption, because with a base value of 1 this process will produce a distribution which has a relatively high probability of zero cases (i.e. no symptomatic acute episodes).

The cost of an episode for these acute hospitalisations is based on the NHS Reference Costs that identify inborn errors of metabolism as a separate Healthcare Resource Group (HRG) entry.[94] The National Schedule of Reference Costs is produced annually by the NHS and provides a summary of average episode costs for a range of individual treatments and procedures. Cost estimates are submitted by all NHS Trusts based on the HRG case-mix classification system.

It is recognised that some of the calculated costs submitted by individual Trusts may still be of relatively poor quality and accuracy. However, the summary tables provided give information on the mean, range and 25th and 75th percentiles of the overall distribution of costs submitted. The Reference Costs tables for 2001 indicate a mean cost of £1043 for non-elective episodes of inborn errors of metabolism. For the purposes of the model, it is assumed these acute hospitalisation costs follow a skewed (log-normal) distribution, with upper and lower limits set according to the range of values reported in the Reference Cost tables. It is considered that the episode costs provided in the NHS Reference Costs provide a conservative estimate of the types of symptomatic acute hospitalisation incurred by MCAD deficiency cases. This is because there are likely to be significant case-mix differences, even within the HRG category for inborn errors of metabolism, related to those cases who present symptomatically (e.g. a significant proportion of presenting patients require intensive care). Thereafter, it was assumed that there would be no systematic difference in the number of healthcare episodes or hospitalisations after diagnosis.[60]

It is also recognised that confirmation of a positive diagnosis would then require appropriate paediatric referral. The NHS cost of this paediatric referral is based on the estimated value of 1 hour of a (medical) consultant's time, published in 'Unit costs of health and social care: 2001.[93] The base cost is therefore £64. It is assumed in the model that a paediatric referral would occur for all cases detected presymptomatically through screening, and only for cases detected symptomatically in the absence of screening.

The third extension was to include some estimate of the future (extra) healthcare and social care costs that may be imposed by some of the more significant disabilities and impairments that may arise from symptomatic presentation. In addition to a range of relatively minor impairments, more debilitating effects can include cerebral palsy. Significant impairments may affect between 10 and 15% of symptomatic cases.[60,92]

Estimates of the future healthcare and social care costs (in a UK context) for these types of impairment are difficult to assess. The economic evaluation of the costs and benefits of the existing UK screening programme for PKU referred to earlier[90] included estimates of future social care costs related to the types of severe disability caused by a failure to detect and treat the condition early. In this evaluation of the UK PKU programme, estimates for future special education, healthcare and social care costs for those with severe disabilities and impairments varied from £168,800 to £278,000 (probabilistic estimate: £252,900), based on 1995 prices.[90] It was therefore decided to accept the probabilistic best estimate derived in this study as indicative of the future healthcare and social care costs associated with the most severe cases of disability caused by MCAD deficiency. The 1995 figure was adjusted to current prices (£290,000) using the Health Services Cost Index.

Nevertheless, it is recognised that there will be a spectrum of disabilities and impairments caused by MCAD deficiency, even across those considered to be significant and likely to impose future costs. It was therefore necessary to obtain an estimate of the future healthcare and social care costs imposed by less severe disabilities than those implied in the PKU study.[90]

With this in mind, a more recent economic assessment of the incremental costs associated with the presence of simple and complex hemiplegic cerebral palsy (HCP)[95] was consulted. This study was designed to identify the extra healthcare and

social care costs imposed by these conditions on young adults (aged 16–24 years) over the course of a year. Although the cost distributions obtained were heavily skewed, the mean incremental costs imposed by HCP were £5600 per person per year, 41% of which were related to 'education and day activities'. This figure was therefore used as an estimate of the future (incremental) costs associated with simple and complex hemiplegic disabilities caused in a proportion of MCAD deficiency cases.

The base-case estimate was derived as follows: annual costs up to 16 years of age were assumed to be the same as those that obtained throughout the young adult years estimated within the study.[95] As service provision appears to decline after childhood (e.g. reductions in physiotherapy, orthotics and orthopaedic support),[95] the incremental costs imposed on young adults may be thought of as a conservative estimate of the extra costs imposed through childhood. The annual estimated cost after 24 years of age was reduced by 41% to remove the 'education and day activities' component. To calculate the future costs associated with this group, the mean future life expectancy of those with HCP was estimated to be around 60–65 years. (With discounting at 6%, a more precise estimate has a negligible effect on the total cost estimate derived.) This is based on research into the life expectancy of cerebral palsy cases in both the UK and the US.[96–98] This procedure produced an estimated lifetime cost associated with hemiplegic disabilities of about £88,000 (2001 prices; discounted at 6%). This is assumed to be relatively conservative.

These two values (£88,000 and £290,000) are taken as minimum and maximum estimates for the range of costs incurred by the occurrence of moderate to severe disability caused by MCAD deficiency. These values are entered into the model as a continuous uniform distribution, to characterise the wide degree of uncertainty on the future costs associated with disabilities caused by MCAD deficiency.

It is recognised that these estimates are relatively crude, and that the future healthcare and social care costs associated with the disability and impairment caused by symptomatic emergence of the condition could have a significant impact on the overall model results. The effects of the uncertainty around these estimates will be examined in more detail as part of the VOI analysis.

In addition to costs, the presence of moderate to severe disability affects available health benefits (expressed in life-years) through a generally lower life expectancy for affected cases. Note, however, that these life expectancy estimates are based on research into cases classified as cerebral palsy in which no reference is made to the presence of additional complications such as an inborn metabolic disorder. Therefore, this estimate of the life expectancy of cases with disabilities due to inborn metabolic disorders based on the experience of cerebral palsy cases alone is an assumption. Furthermore, although the conditions between the USA and UK may not be comparable, this information has only been used to approximate life expectancy for individuals affected by moderate to severe disabilities and impairments. Study data obtained from both the UK and the USA have been used to estimate the wide variation in life expectancy according to different types of impairment and severity.[97–100] For the most severe forms of disability a life expectancy of 35 years is assumed, for more moderate cases a life expectancy of 55 years is assumed, whereas for cases with mild impairments a near normal life expectancy, of 65 years, is assumed. The uncertainty around these estimates of life expectancy and the relative proportions affected by moderate to severe forms is characterised by adopting an asymmetric triangular distribution (35, 55, 65), where 55 represents the most probable value. Asymptomatic cases and those without significant impairments are assumed to have normal life expectancy based on current (2001) UK life-table calculations (approximately 75 years for males, 80 years for females).

As reported in the clinical review (Chapter 6; MCAD deficiency, 'Discussion', p. 55), various studies of infants with MCAD deficiency have found that, following early diagnosis and initiation of effective treatment, no deaths have occurred[20,23,28,63] and there have been no cases of appreciable cognitive impairment or neurological damage.[28,63] Therefore, the economic model assumes a value of zero for the proportion of screened cases who die or develop significant disabilities or impairments.

All parameters and distribution assumptions used in the model for PKU and MCAD deficiency combined are summarised in Appendix 40.

GAI and homocystinuria (due to CβS deficiency)
As indicated previously, additional information of relevance to the economic modelling was available in the literature for both GAI and homocystinuria (due to CβS deficiency). Therefore, modelling the

impact of neonatal screening using tandem MS was extended to include each of these conditions in more detail.

GAI The incidence of GAI in the UK was taken from the estimate of 2 per 100,000 used in the previous HTA report, which the authors suggest is an underestimate.[1] The incidence per cohort of 100,000 neonates was characterised in the model using a Poisson distribution. Similarly to MCAD deficiency, it is recognised that a proportion of those detected through screening will include some cases that would have remained asymptomatic in the absence of screening (base-case estimate: 20%[1]).

Information on the sensitivity and specificity of screening was obtained from the literature reviewed on the accuracy of tandem MS for acylcarnitine disorders. Thus, sensitivity appeared to be 100%, while the false-positive rate was estimated to be 0.023 (95% CI 0.0159 to 0.0297%). A false-negative rate of zero is assumed for a relatively rare condition such as this, and it is recognised that this may be an optimistic assumption.

Treatment profiles and costs, for those detected presymptomatically and those detected symptomatically, were based on those prescribed for GAI in the previous HTA economic model[1] and described in Appendices 34–36. The costs applied to screened and unscreened cohorts were characterised as a continuous uniform distribution between the estimated base-case value and twice the base-case value.

The following assumptions used in the MCAD deficiency model are common to the GAI model.

- A further £0.10 per sample was added to account for additional reagent costs, internal standards and other consumables for the extension of screening to an additional disorder.
- Symptomatic presentation will impose the cost of at least one acute hospitalisation episode at a mean cost of £1403 (with a log-normal distribution).
- Confirmation protocols are required for both presymptomatic (£130) and symptomatic cases (£150), but there will be systematic differences based on the number of false positives generated through screening and the proportion of cases that remain asymptomatic in the absence of screening.
- Confirmation of a positive diagnosis is attended by an appropriate paediatric referral (with an estimated base cost of £64).
- The same cost limits used within the MCAD deficiency model (£88,000 and £290,000) are taken as minimum and maximum estimates for the range of costs incurred by the presence of moderate to severe disability caused by GAI deficiency.
- Estimates of life expectancy for the spectrum of disabilities and impairments that affect GAI cases are assumed to be the same as for MCAD deficiency. Life expectancy for asymptomatic cases and those who do not develop significant disabilities is assumed to be near normal (65 years).

One key difference from the MCAD deficiency model is that the incremental cost of using tandem MS added in this model is based on including the extra costs (positive or negative) obtained. MCAD deficiency has been added to the PKU programme.

Parameter values and the distribution assumptions used in the model for GAI are summarised in Appendix 41.

Homocystinuria (due to CβS deficiency) The incidence of homocystinuria (due to CβS deficiency) in the UK was taken from the estimate of 1.5 per 100,000 used in the previous HTA report.[1] The incidence per cohort of 100,000 neonates was characterised in the model using a Poisson distribution.

Information on the sensitivity and specificity of screening for homocystinuria (CβS) using tandem MS was limited. Data from the New England Newborn Screening program, USA, found a false-positive rate of 0.0272% (95% CI 0.0209% to 0.0336%).[7] However, this study did not report the sensitivity of neonatal screening for homocystinuria (based on the detection of high methionine levels); therefore, a false-negative rate of zero was assumed.

Treatment profiles and costs, for those detected through screening and those detected symptomatically, were based on those prescribed for homocystinuria as set out in the previous HTA economic model[1] and described in Appendices 34–36. The costs applied to screened and unscreened cohorts were characterised as a continuous uniform distribution between the estimated base-case value and twice the base-case value.

The following assumptions used for the evaluation of MCAD deficiency were applied within the model for homocystinuria.

- A further £0.10 per sample was added to account for additional reagent costs, internal standards and other consumables for the extension of screening to an additional disorder.
- Symptomatic presentation will impose the cost of at least one acute hospitalisation episode at a mean cost of £1403 (with a log-normal distribution).
- Confirmation protocols are required for both presymptomatic and symptomatic cases identified (at a cost of £160). Overall confirmation costs for the screening option will be higher because false positives will be generated through screening.
- Confirmation of a positive diagnosis is attended by an appropriate paediatric referral (with an estimated base cost of £64).
- The same cost limits used within the MCAD deficiency model (£88,000 and £290,000) are taken as minimum and maximum estimates for the range of costs incurred by the presence of moderate to severe disability caused by homocystinuria.
- Estimates of life expectancy for the spectrum of disabilities and impairments, which affect homocystinuria cases, are assumed to be the same as for MCAD deficiency. Life expectancy for cases detected and compliant with treatment is assumed to be near normal (65 years).

A number of additional specific assumptions used within this model are worth highlighting.

- All cases will become symptomatic at some stage if not detected early through a screening programme.
- Not all cases of homocystinuria (CβS) detected through a screening programme will be compliant with treatment and therefore avoid the complications associated with this condition. In one study of a screening programme for homocystinuria, four individuals developed mental disability (two detected through screening but non-compliant with treatment).[48] In a further study, 32% of those detected by screening were non-compliant and subsequently developed complications.[49] The model, therefore, assumes that between 10 and 15% of cases detected through screening will develop significant disabilities as a consequence of poor treatment compliance. However, in the absence of screening 60% of all homocystinuria cases will develop significant disabilities.[51]
- In the absence of screening, 22.5% of cases will die (with a mean age at death of 12.9 years).[47]

As with the GAI model, any incremental cost of using tandem MS is added into the homocystinuria model using the calculated extra costs (positive or negative) obtained after MCAD deficiency has been added to the PKU programme.

Parameter values and the distribution assumptions used in the model for homocystinuria (CβS) are summarised in Appendix 42.

Results

PKU screening: existing technologies and tandem MS

Results for phase i were derived from a number of Monte Carlo simulations (using 10,000 iterations) using model parameters and associated distributions as set out in Appendix 39 and explained in detail above. Given the potential scale economies possible from a highly capital-intensive technology such as tandem MS, it is clear that the distribution of incremental costs associated with the adoption of tandem MS for the national screening programme for PKU depends on the volume of specimens processed on a single system. This is illustrated in *Table 12*, which shows the mean incremental costs obtained for different hypothetical volumes for a given tandem MS system.

Table 12 shows that based on the available evidence used in the economic model, at volumes below 50,000 specimens per year the use of tandem MS for the current PKU programme would incur significant additional costs. As expected, at much higher processing volumes (around 80,000) the incremental cost of using tandem MS for neonatal screening of PKU is far less, but still not negative.

To test the model results based on what might be considered a realistic scale of operations for participating laboratories, the analysis was also performed based on the cost per specimen obtained when processing volumes varied (as a continuous uniform distribution) between 50,000 and 60,000 samples per year. The results from the simulation exercise for the range between 50,000 and 60,000 specimens is also given in *Table 12* and the distribution is illustrated in *Figure 4*.

The mean and standard deviation derived from this sampling distribution were £54,900 and £15,899, respectively. The size of the standard deviation on this distribution of incremental costs indicates the degree of variation present. A comparison of the results for tandem MS against the existing technologies (based on the range of 50,000–60,000 samples per annum for tandem MS) is presented in *Table 13*. Note that the

TABLE 12 *Cost comparison: PKU screening using existing technologies or tandem MS for different volumes of tandem MS use (for a cohort of 100,000 neonates)*

Volume of specimens per tandem MS system	Incremental cost		*p*(Incremental cost < 0)
	Mean	SD	
20,000	£308,929	£26,056	0.0000
40,000	£110,967	£16,790	0.0000
50,000	£71,295	£15,176	0.0000
60,000	£45,127	£14,501	0.0009
70,000	£26,622	£13,723	0.0255
80,000	£12,731	£13,385	0.1729
Varies from 50,000 to 60,000 per system	£54,900	£15,899	0.0004

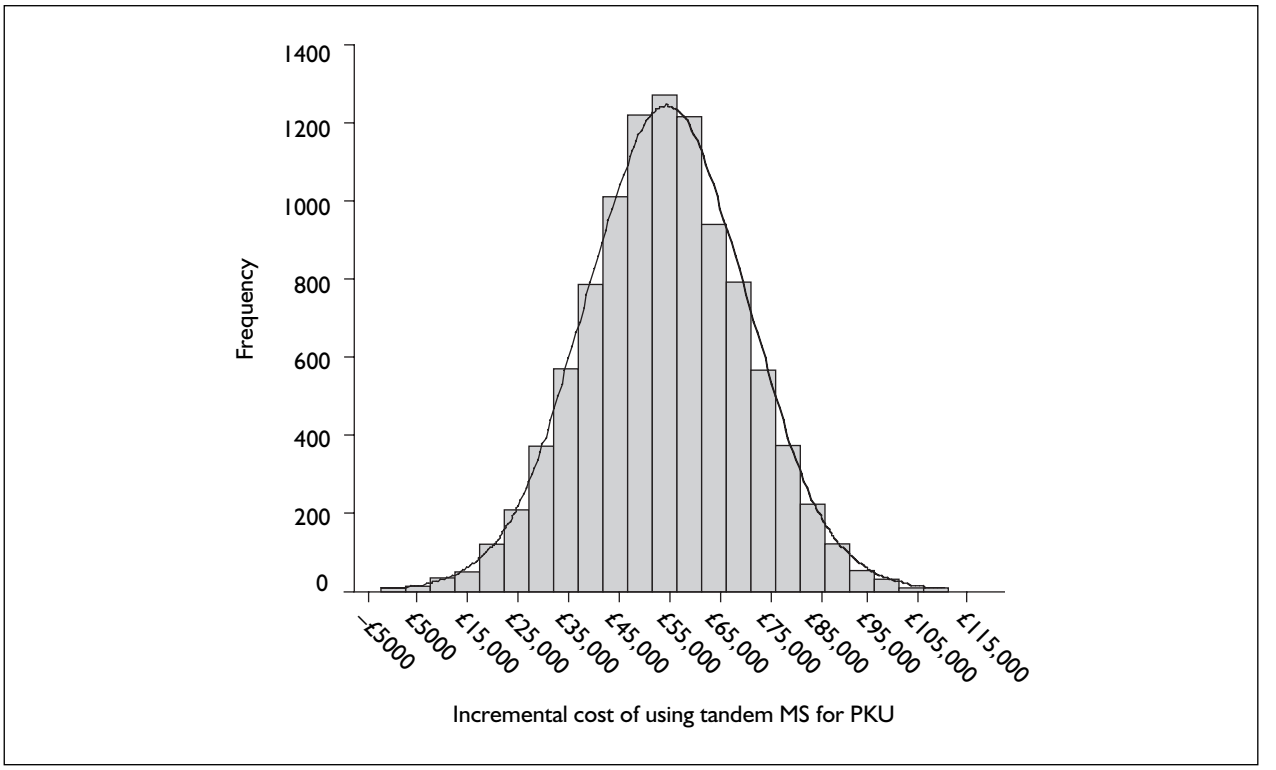

FIGURE 4 *Incremental cost of adopting tandem MS for PKU screening (per 100,000) based on a range of volumes between 50,000 and 60,000 specimens per system*

TABLE 13 *Cost comparison: PKU screening using existing technologies or tandem MS*

		(1) Existing	(2) Tandem MS	Incremental cost (2)–(1)
Total cost of screening[a]	Mean	£114,528	£169,428	£54,900
(for 100,000 neonates)	SD	£12,468	£10,597	£15,899
Cost per neonate screened		£1.15	£1.69	£0.55
Cost per case detected		£12,725.34	£18,825.34	£6,100

[a] Comprises all laboratory costs (including advice and referral), all repeat sampling and confirmation costs. Excludes the cost of the initial specimens collected.
Note: Estimates of total and incremental costs are for a cohort of 100,000 neonates screened.

estimates of total and incremental costs are for a cohort of 100,000 neonates screened.

Tables 12 and *13*, and the distribution of incremental costs illustrated in *Figure 4*, indicate that the use of tandem MS for the current PKU screening programme in the UK would not be cost-effective.

Priority ordering for additional conditions

Phase ii included a straightforward addition of the mean incremental cost of using tandem MS for PKU screening (*Table 13*) into each decision model for additional conditions detectable by tandem MS. The priority ordering derived in this way, therefore, differed from that obtained in the earlier HTA modelling exercise.[1] The latter was based only on a comparison of the incremental costs of treatment with the incremental life-years gained, without including an allowance for any additional burden of cost associated with the adoption of the more expensive technology (tandem MS) for the existing PKU screening.

The results obtained using this simple sequential analysis are presented in *Table 14*. This shows that – based on the frequency of the condition, the costs of treatment incurred or future costs avoided, the potential health effects and additional life-years gained – the first priority would be to extend any neonatal screening to include MCAD deficiency.

Too much should not be read into the priority list presented in *Table 14*. This table represents an ordinal ranking only; a number of the conditions produced similar results (although MCAD deficiency clearly produced the most significant gains when added to the PKU programme using tandem MS).

TABLE 14 *Priority ordering of inborn metabolic conditions for extended screening using tandem MS*

Condition	Rank
MCAD deficiency	1
GAI	2
Long-chain fatty acid defects	3
Homocystinuria	4
Urea cycle disorders	5
GAII	6
Methymalmonic acidaemia	7
Branched-chain acyl-CoA metabolism defects	8
Propionic acidaemia	9
Isovaleric acidaemia	10
Tyrosinaemia (type 1)	11
MSUD	12

It should be emphasised that this initial ranking is based on estimating the cost per life-year gained when the full additional cost of using tandem MS for PKU is included with the other costs and compared with the life-years gained for each condition separately. The same ordinal ranking would not be obtained when the extra cost of using tandem MS had been offset by extension to the first condition to be included with PKU screening. Then, the incremental costs of screening for additional disorders would be far smaller and the ranking would change to reflect mainly differences in treatment costs and life-years gained.

One further issue, however, that affects this initial ranking of conditions is the inclusion of an estimate of cost savings associated with the avoidance of moderate to severe disabilities. It has been possible to estimate this for MCAD deficiency, GAI and homocystinuria (CβS) using published sources that indicate the proportion of symptomatic cases who are likely to avoid disability/impairment through screening and early initiation of effective treatment. Unfortunately, there is far less certainty, at least from the available published sources to date, about what fraction of symptomatic cases would be affected by moderate to severe disability/impairment for other conditions. Consequently, estimates of future costs related to avoidable disability and impairment for these other inborn metabolic conditions were not included.

It should be emphasised, however, that the purpose of this ranking exercise was to establish which condition would provide the most impact when added to a PKU screening programme based on tandem MS technology.

Extending PKU screening with tandem MS

MCAD deficiency

In many conventional cost-effectiveness studies, a simple deterministic worst and best case scenario is constructed using the least and most favourable input values to indicate the limits or extreme results that are possible. This is difficult in the current context for two reasons. First, it is important to see how the CEA results vary according to the different processing volumes (operating scales) evaluated, so that the implications for service delivery and organisation can be assessed appropriately. Second, some important model parameters have wide margins of variation because of the uncertainty surrounding available knowledge or information.

In such circumstances, modelling a best and a worst case scenario would produce such disparate

TABLE 15 *Cost of using tandem MS for PKU plus MCAD deficiency screening compared with existing PKU only*

	(1) Existing (PKU only)	(2) MS-MS (PKU+MCAD deficiency)	Incremental cost (2)–(1)
Total cost of screening for 100,000 neonates (mean)	£114,528	£91,216	–£23,312
Cost per neonate screened	£1.15	£0.91	–£0.24
Cost per case of inborn error of metabolism detected	£12,725	£5,366	–£7,359

and asymmetrical results that they would have limited practical value in the interpretation of the cost-effectiveness results. The advantage of the probabilistic modelling using the Monte Carlo simulation approach used here is that iterations sample across the range of possible parameter distributions. This is more appropriate because the underlying uncertainty, characterised by the input values and distributions used, affects the overall outcome more realistically than an artificial best and worst case analysis.

Results from the Monte Carlo simulation modelling indicate that the addition of screening for MCAD deficiency as part of a neonatal screening programme for PKU using tandem MS would be cost-effective.

Using an operational range of 50,000–60,000 specimens per system per year, the mean incremental cost for PKU plus MCAD deficiency screening using tandem MS from the model was –£23,312 (median: –£14,810); this is after allowing for the incremental cost incurred when using tandem MS for PKU only. The implications of this are presented in *Table 15*. This cost saving is also associated with a mean incremental gain of 59 life-years for each cohort of 100,000 neonates screened.

Note that the tandem MS estimates represent the impact of additional costs associated with the use of the new technology for the PKU programme and costs related to MCAD deficiency screening, net of any savings attributable to avoidable future disability costs and other avoidable costs (such as unnecessary symptomatic episodes) that would occur in the absence of screening.

Although *Table 15* provides information based on the mean value derived from the simulation exercise, it is important to recognise the nature of the distribution obtained. *Figure 5* illustrates that the distribution of incremental costs is negatively skewed (the 5th and 95th percentiles range from –£142,489 to £62,350). Intuitively, this distribution of costs obtained (based on 10,000 iterations of the model) makes sense, reflecting the impact of the potential future healthcare and social care costs related to disabilities avoided by screening. On some cohorts of 100,000 samples screening may not detect any cases or only those in the relatively mild or asymptomatic spectrum of the disease and this will incur additional (net) costs. On other occasions screening will allow early identification and treatment of cases who would have developed severe disabilities. The future healthcare and social care costs of a relatively few cases with severe disabilities caused by MCAD deficiency would be large, and these can be avoided by screening.

The mean incremental gain in life-years (which is independent of the operating scale of tandem MS) is 59. Both the incremental costs and life-year gains reported above are for a cohort of 100,000 neonates screened.

The simulation results are illustrated in *Figure 6* using a diagram that shows the distribution of the incremental cost per-life year gained derived and presented on a conventional cost-effectiveness plane. Again, this is based on input values related to an operational range of 50,000–60,000 specimens per year. Note that preference is given to the presentation of results using CEAcc rather than a simple cost per life-year gained because of the potential ambiguities of interpretation with a negative ICER. A range of CEAcc is plotted in *Figure 7*, each representing the results obtained for different operating volumes of a tandem MS system.

The CEAcc diagram suggests that, on the basis of the available evidence applied in this model, even at relatively low threshold values (i.e. willingness to pay for an additional life-year gained) there is a high probability that using tandem MS for PKU and MCAD deficiency screening combined would be cost-effective. As might be expected, the probability of cost-effectiveness increases at higher operating volumes of a tandem MS system. At a

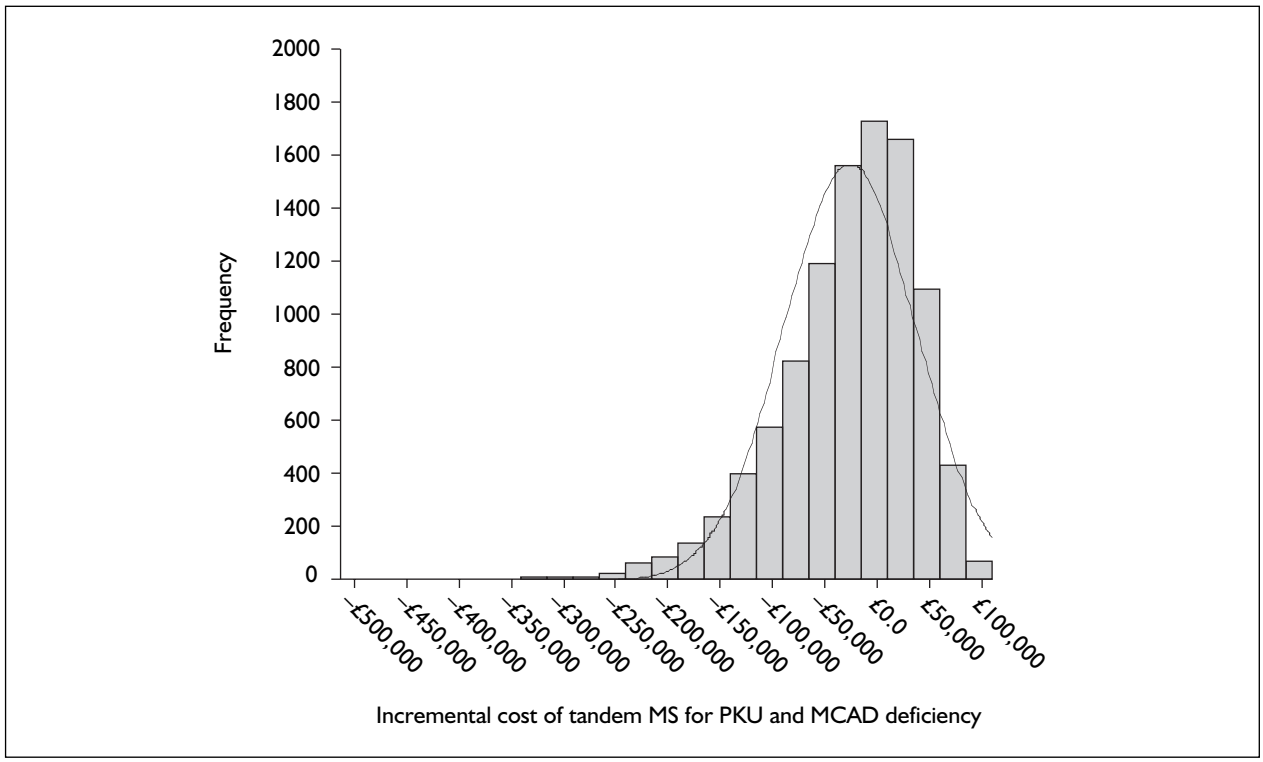

FIGURE 5 *Incremental cost of extension of the PKU programme to include screening for MCAD deficiency using tandem MS*

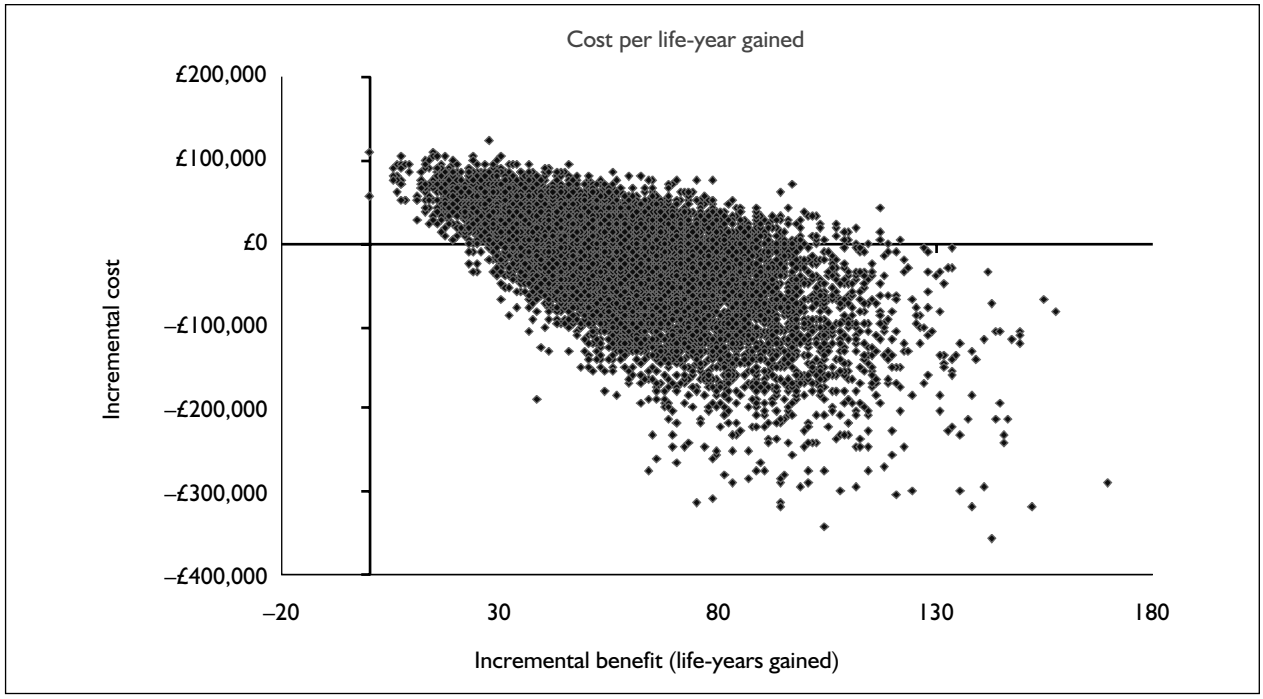

FIGURE 6 *Cost-effectiveness plane: PKU plus MCAD deficiency screening with tandem MS*

very low volume (20,000 samples per year) the probability that the use of tandem MS would be cost saving is zero. At volumes of between 50,000 and 60,000 samples the economic model indicates that the use of tandem MS would produce modest cost savings for each cohort of 100,000 screened. Moreover, for 50,000–60,000 samples per year, at a threshold value of £1000 the probability that using tandem MS is cost-effective is 0.86, and at a value of £5000 per life-year gained the probability

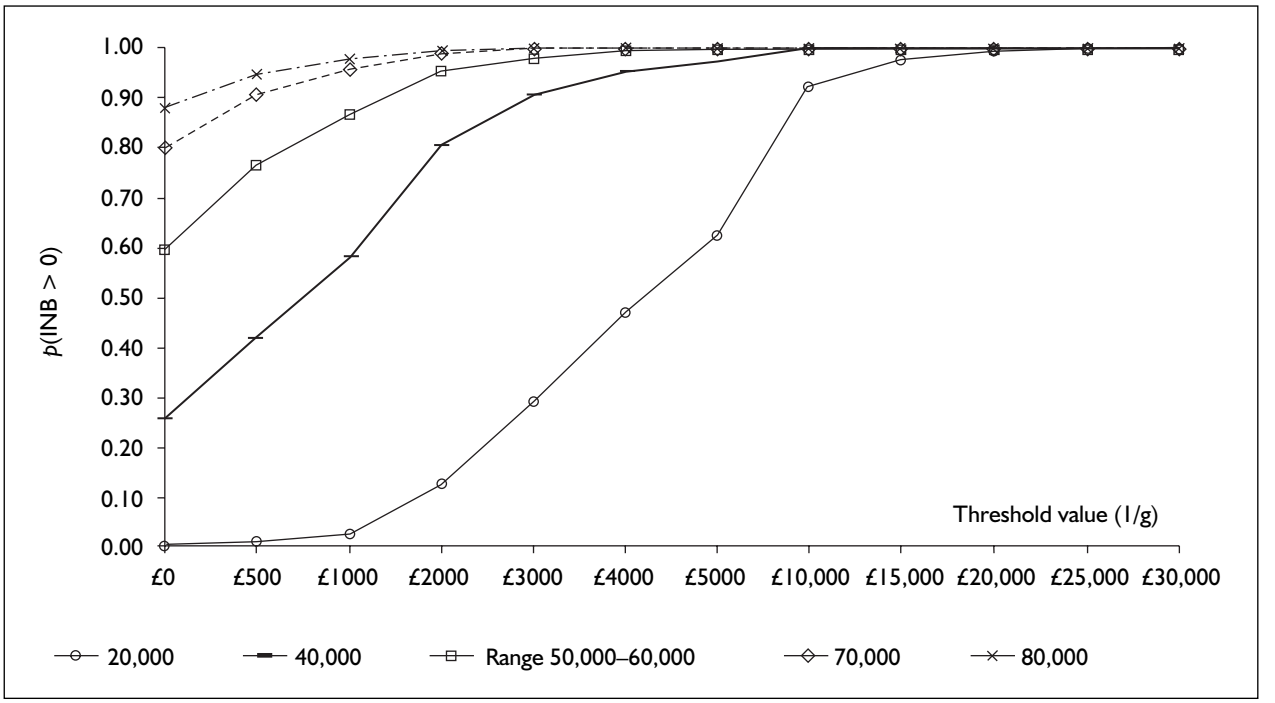

FIGURE 7 *CEAcc for PKU plus MCAD deficiency screening using tandem MS.*
Note: the horizontal axis is not in a uniform interval scale

is 0.99. As indicated previously, this represents a realistic operating range for a single tandem MS system in a screening programme.

GAI and homocystinuria (due to CβS deficiency)
An important issue for any further development of this economic model for PKU and MCAD deficiency using tandem MS is whether any extension of the analysis should be evaluated in the context of a neonatal screening package. In other words, should any estimated net savings obtained through extending neonatal screening for one metabolic disorder then be carried over into the model for the next priority condition identified as part of the stage ii analysis? The economic implications of this are examined using probabilistic modelling of extending neonatal screening to include GAI and homocystinuria (CβS).

Using the incremental cost results and life-year gains obtained when using tandem MS screening for PKU and MCAD as input parameters, further Monte Carlo simulation models were constructed to evaluate the effects of using tandem MS to include screening for GAI and homocystinuria (CβS). Each was assessed independently (that is, the results obtained from the PKU plus MCAD deficiency model were used as inputs into the models for GAI and homocystinuria separately). However, because the results obtained for PKU plus MCAD deficiency were highly skewed it would be inappropriate to enter the cost results obtained from this model as a normally distributed variable. The incremental costs effects of screening for PKU plus MCAD deficiency using tandem MS were therefore entered into both models using a triangular distribution based on 5th, mean (as most probable) and 95th percentile values reported earlier. This will tend to understate the incremental savings obtained from MCAD deficiency screening by excluding a proportion of very large negative values from the overall distribution illustrated in *Figure 5*. All input parameters and distributions for GAI are set out in Appendix 41, while all input parameters and distributions for homocystinuria (CβS) are set out in Appendix 42. Both economic models are based on an assumed operating range of 50,000–60,000 specimens per year for a tandem MS system.

GAI The mean incremental cost for PKU plus MCAD plus GAI screening using tandem MS from the Monte Carlo simulations was £23,614 (median: £22,268). This is for a cohort of 100,000 neonates screened. An interesting feature of these results obtained for GAI is that overall costs may actually increase, offsetting the mean cost savings obtained from the use of tandem MS for PKU and MCAD deficiency combined. This is largely due to a small net increase in future costs associated with disabilities and impairments as a consequence of screening. This occurs because some individuals

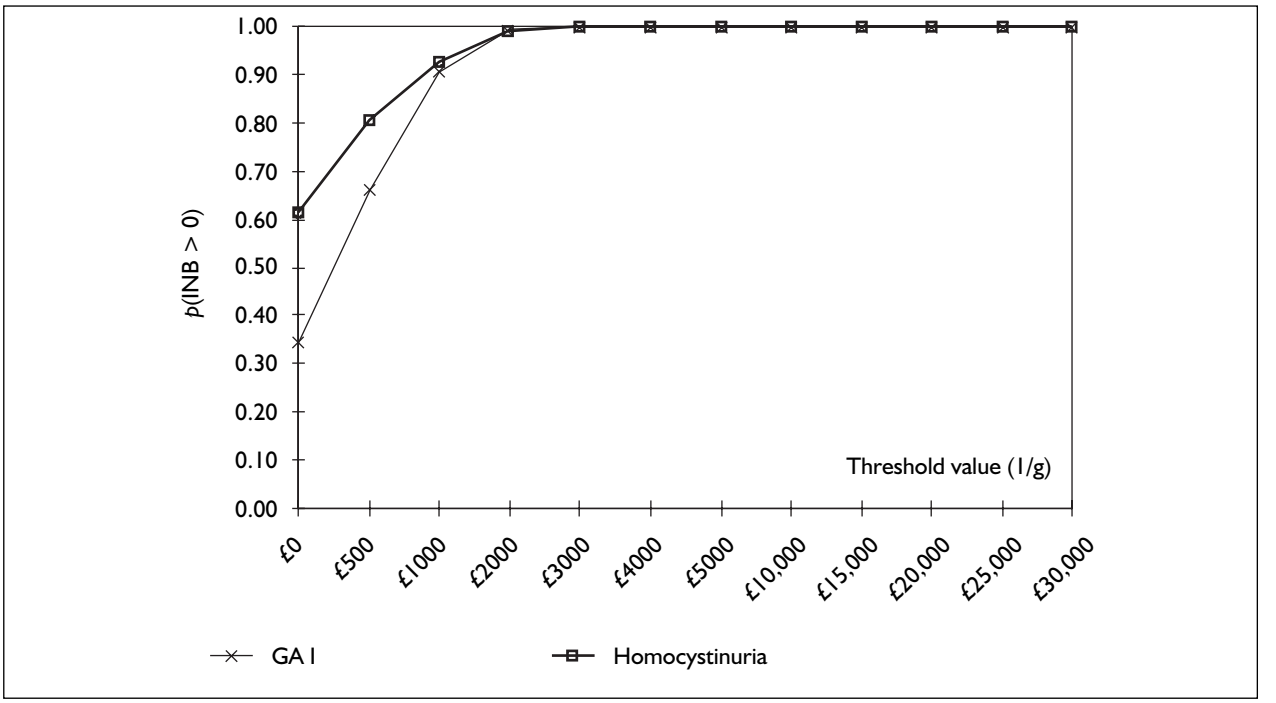

FIGURE 8 *CEAcc for combined PKU and MCAD with additional disorders identified using tandem MS. Note: the horizontal axis is not in a uniform interval scale*

identified through screening will still suffer disabilities. Although screening will significantly reduce early mortality from GAI, this means that an additional number of cases will survive and become susceptible to subsequent complications that cause disabilities.[69]

The mean incremental life-years gained for both MCAD deficiency and GAI screening combined using tandem MS technology was 90 for each cohort of 100,000 neonates, equivalent to a cost per life-year saved of £261. On the basis of the available evidence used in this model, the simulation results indicate a high probability that the use of tandem MS for the current PKU programme, with extension to include screening for MCAD deficiency and GAI, would be cost-effective at relatively low threshold values. At a threshold value of zero the likelihood that this option is cost-effective is only 0.34; however, at a modest threshold value of £1000 the probability is 0.91. This can be seen clearly in *Figure 8*, which shows the cost-effectiveness acceptability curve for different threshold values for GAI. The distribution of cost per life-year values obtained from the probabilistic model is illustrated in the cost-effectiveness plane in *Figure 9*.

Homocystinuria (due to CβS deficiency) The mean incremental cost for PKU plus MCAD plus CβS) screening using tandem MS from the Monte Carlo simulations was –£26,833 (median: –£19,669) for a cohort of 100,000 neonates screened. Extension of neonatal screening using tandem MS for PKU, MCAD deficiency and homocystinuria combined produces a similar overall cost saving to that obtained by PKU and MCAD deficiency alone. In other words, any extra variable costs incurred through the addition of homocystinuria, including treatment costs, are offset by savings that arise from reduced disabilities and impairments.

The mean incremental life-years gained for both MCAD deficiency and homocystinuria (CβS) screening combined using tandem MS technology was 74 for each cohort of 100,000 neonates, equivalent to a cost per life-year saved of –£360. The simulation results indicate a high probability that the use of tandem MS for the current PKU programme, with extension to include screening for MCAD deficiency and homocystinuria (CβS), would be cost-effective at low threshold values. Based on the available evidence used in the economic model, at a threshold value of zero the likelihood that this option is cost-effective is 0.62, and at a threshold value of £1000 the probability is 0.93. This can be seen in *Figure 8*, which shows the CEAcc for different threshold values for homocystinuria. The distribution of cost per life-year values obtained from the probabilistic model is illustrated in the cost-effectiveness plane in *Figure 10*.

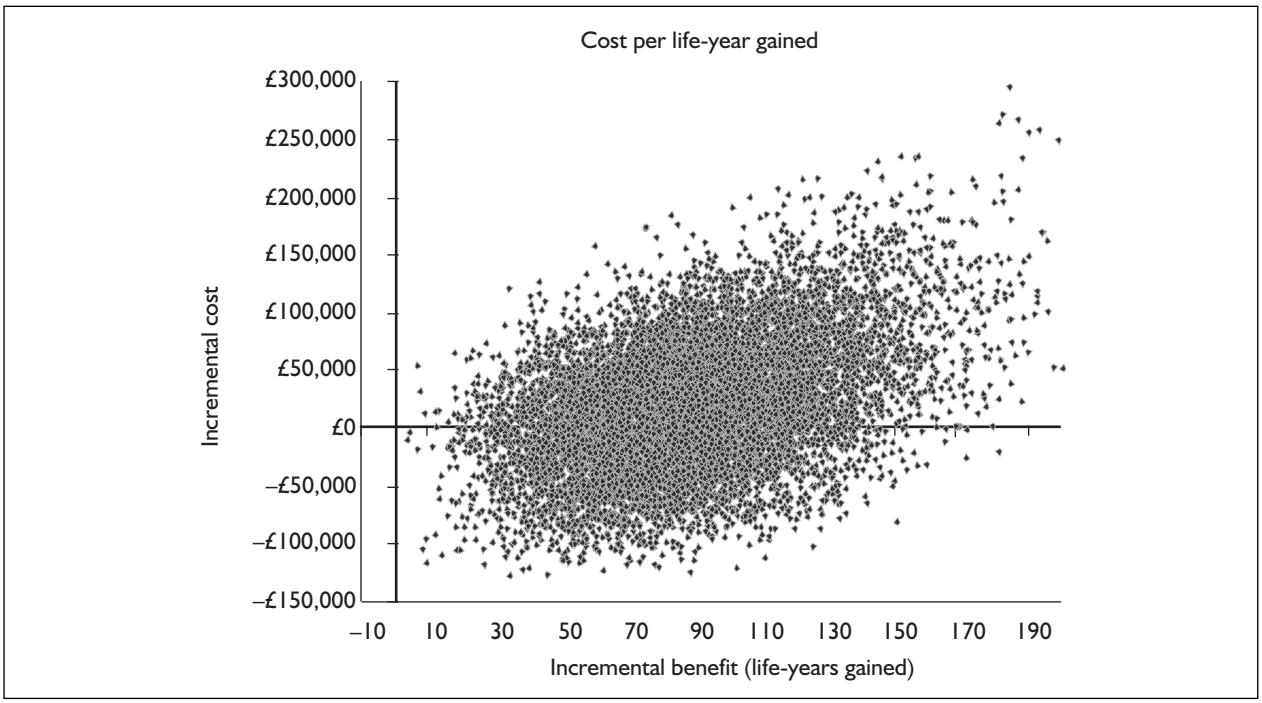

FIGURE 9 *Cost-effectiveness plane: PKU plus MCAD plus GAI screening with tandem MS*

Other inborn errors of metabolism

Having established that the use of tandem MS for PKU and MCAD deficiency screening would, at the very least, offset any additional costs related to use of the newer technology, simulations were also run to produce simple cost per life-years gained and INB estimates for each of the other main conditions. These models were constructed using the assumptions adopted in phase ii, but now including the incremental cost saving associated with adopting tandem MS for PKU and MCAD deficiency screening combined. Note, however, that these models did assume some additional costs in neonatal laboratories for extended screening and included all other related variable costs (confirmation protocols, paediatric referral, etc.). The assumptions used were set out earlier.

To provide a consistent basis on which to compare the results obtained for GAI and homocystinuria, each of the other inborn metabolic disorders was modelled separately. As very little additional material of relevance to the economic assessment of these conditions has been published since the two earlier systematic reviews,[1,2] most of the parameter assumptions used are based on the information set out within the previous economic modelling.[1]

The results for each condition in terms of the mean cost per life-year gained are shown in *Table 16*.

Of the remaining inborn metabolic disorders considered, screening for long-chain fatty acid defects produces the lowest cost per life-year gained. All but MSUD and tyrosinaemia type I incur an additional cost per life-year saved of less than £6000.

The CEAcc for each of these conditions, for different threshold values, are shown in *Figure 11*. Many of the conditions achieve a high probability of cost-effectiveness (0.8 or above) at a threshold value of £20,000. This is based on the available evidence used in the models as described. However, these results could significantly underestimate the cost-effectiveness of neonatal screening for these disorders using tandem MS. A major limitation of these results is that no information was available in the literature to allow estimates of potential cost savings attributable to avoidable disabilities. Many of these conditions are also responsible for severe disabilities and impairments, and the inclusion of such estimates could significantly affect the relative cost-effectiveness results obtained.

The lack of detailed research information and therefore the levels of uncertainty around long-term outcomes and economic cost of many of these conditions is the primary reason why more detailed consideration was not given assessing the potential cost-effectiveness of extending screening

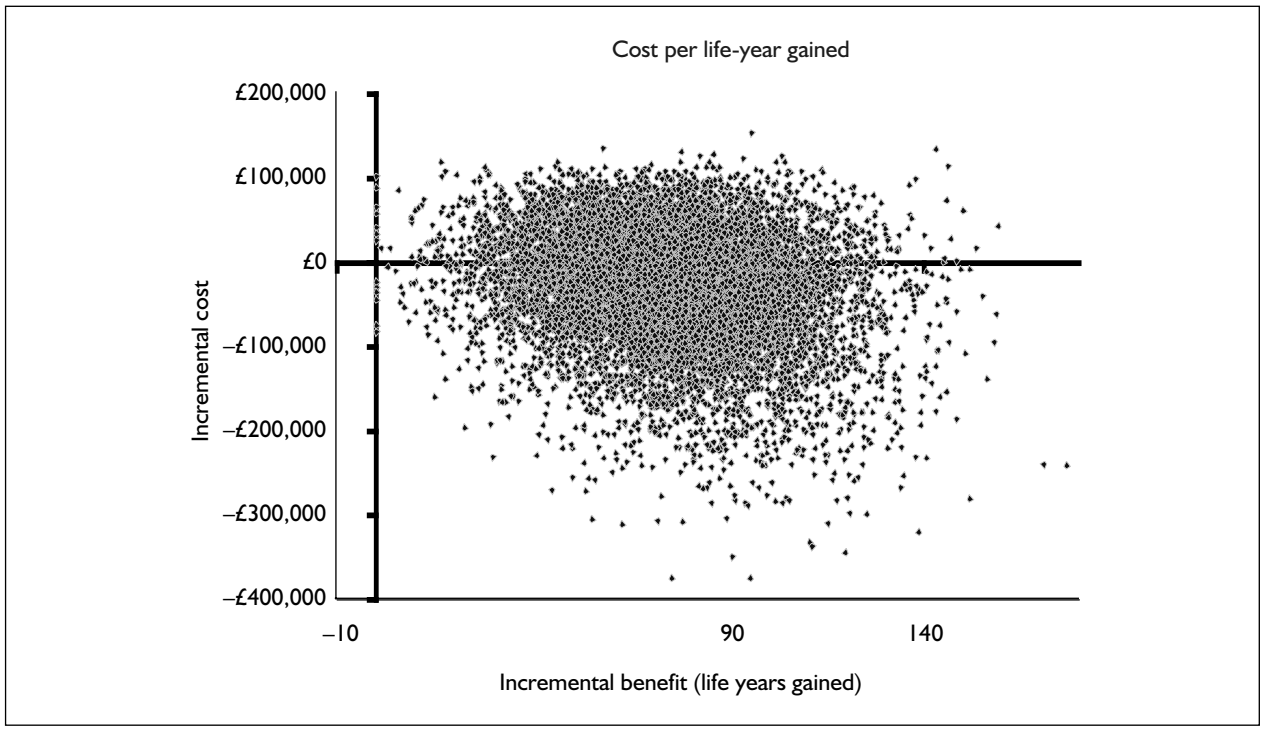

FIGURE 10 *Cost-effectiveness plane: PKU plus MCAD plus homocystinuria (CβS) screening with tandem MS*

TABLE 16 *Cost per life-year gained for other inborn metabolic disorders detectable by tandem MS*

Condition	Cost per life-year gained
Long-chain fatty acid defects	£992
GAII	£2,245
Urea cycle disorders	£2,965
Branched-chain acyl-CoA metabolism defects	£3,694
Methymalmonic acidaemia	£4,638
Isovaleric acidaemia	£5,422
Propionic acidaemia	£5,778
MSUD	£12,388
Tyrosinaemia	£13,168

to these disorders at this stage. This should be the subject of future research.

The value of additional research

The robustness of the results presented above depends on an acceptance of the parameter estimates and distributions used in the model. All model parameters are attended by uncertainty. One advantage of the Monte Carlo simulation modelling undertaken as part of this review is that the work was performed within an EVPI framework. This approach enables decision-makers to estimate the value of additional information about inputs used within the cost-effectiveness model to establish whether the degree of uncertainty within the results warrants further research, and to indicate where this research should be focused.

As most of the uncertainty about the possible application of tandem MS relates to whether any extension of screening to include MCAD deficiency could justify the additional cost, the results presented here are based on the value of additional information about neonatal screening for MCAD deficiency using tandem MS.

The first and most striking result of the EVPI analysis is the extent to which the overall value of additional information appears limited. This is not entirely unexpected, perhaps, given the high probability of cost-effectiveness even at relatively low threshold values. The population expected value of perfect information (PEVPI) for all model parameters at different thresholds is shown in *Table 17*. This is based on the results obtained from the modelling of data for a cohort of 100,000 neonates, adjusted to reflect the value to the total UK neonatal screening population (700,000 neonates per annum). In addition, it is assumed that any additional research information would be of value for a period of 5 years (future costs discounted at 6%). These results are calculated using an assumed operational capacity of between 50,000 and 60,000 specimens per tandem MS system.

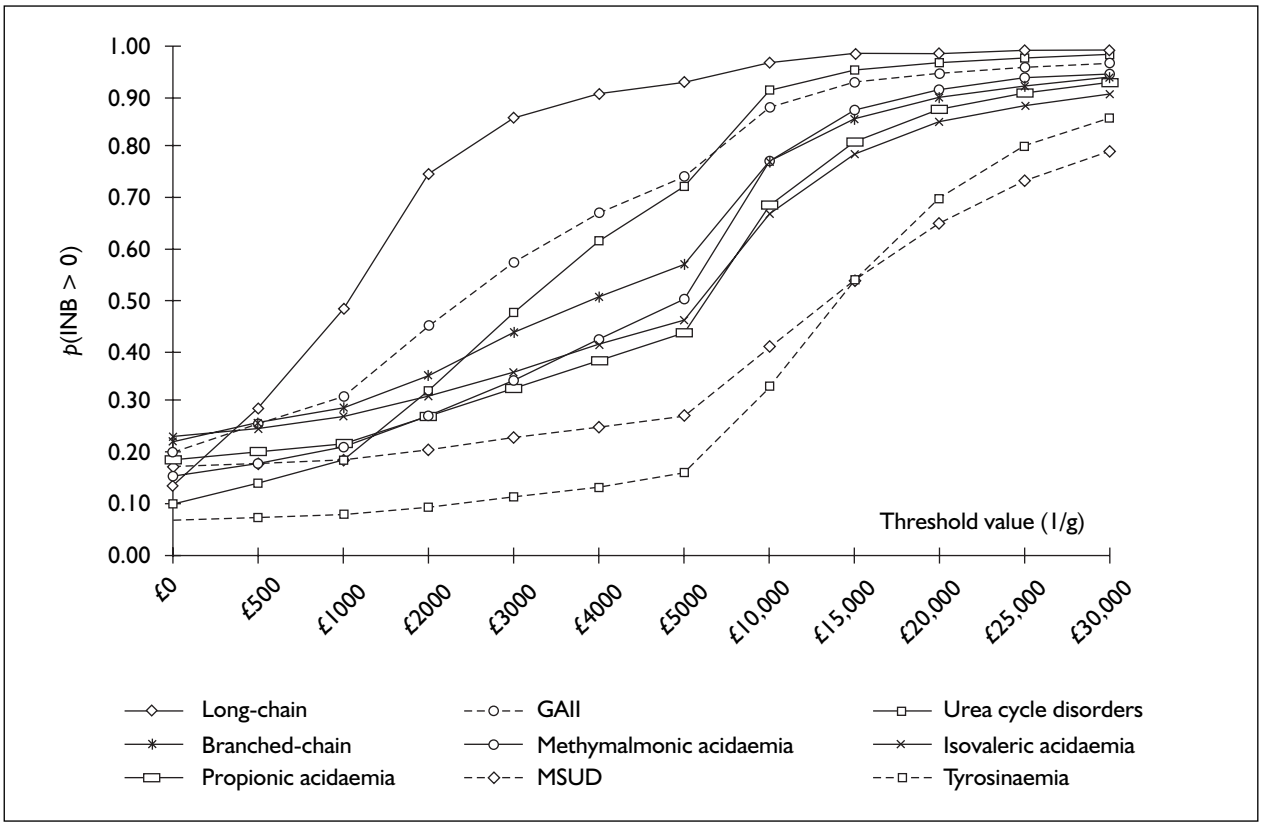

FIGURE 11 *CEAcc for other inborn metabolic disorders detectable by tandem MS.*
Note: the horizontal axis is not in a uniform interval scale

TABLE 17 *PEVPI for all parameters at different threshold values*

Threshold	EVPI
0	£343,144
500	£91,416
1000	£46,832
2000	£3,656
5000	£0

Certainly, at threshold values of £2000 per life-year gained and above, there would be little additional value from further research into whether using tandem MS for PKU and MCAD deficiency screening would be cost-effective. At such threshold values it is effectively certain that using tandem MS for PKU and MCAD deficiency screening is cost-effective (based on the available evidence used in the model).

Only at a threshold value of £500 or less would there be any really substantial gains from further research. *Figure 12* shows the EVPI for specific model parameters at a threshold value of zero.

The results shown in *Figure 12* show the expected value of obtaining perfect information for all and individual parameters within the decision model. This approach enables the identification of those parameters for which further data collection would result in the greatest pay-off in terms of expected net benefit. Partial EVPI for parameters essentially represents an upper bound on the value of conducting further research on a given parameter of interest (θ_I). Partial EVPI analysis for individual or subsets of parameters requires a two-level algorithm, which uses two nested levels of Monte Carlo sampling over the plausible ranges for both parameters of interest, and the remaining uncertain parameters. This process involves sampling the parameter of interest once from its prior uncertain range and holding it at its sampled value, while allowing all parameters not of interest to vary according to their prior uncertainty over 1000 samples. The next stage involves calculating the optimal decision strategy, that is the strategy that achieves the greatest expected net benefit, given the new knowledge on the parameter of interest. This process is then repeated over a large number of iterations to sample over the full range of θ_I, and the mean net benefits are then calculated. The expected value of gaining perfect information on the parameter of interest is thus calculated as the average expected

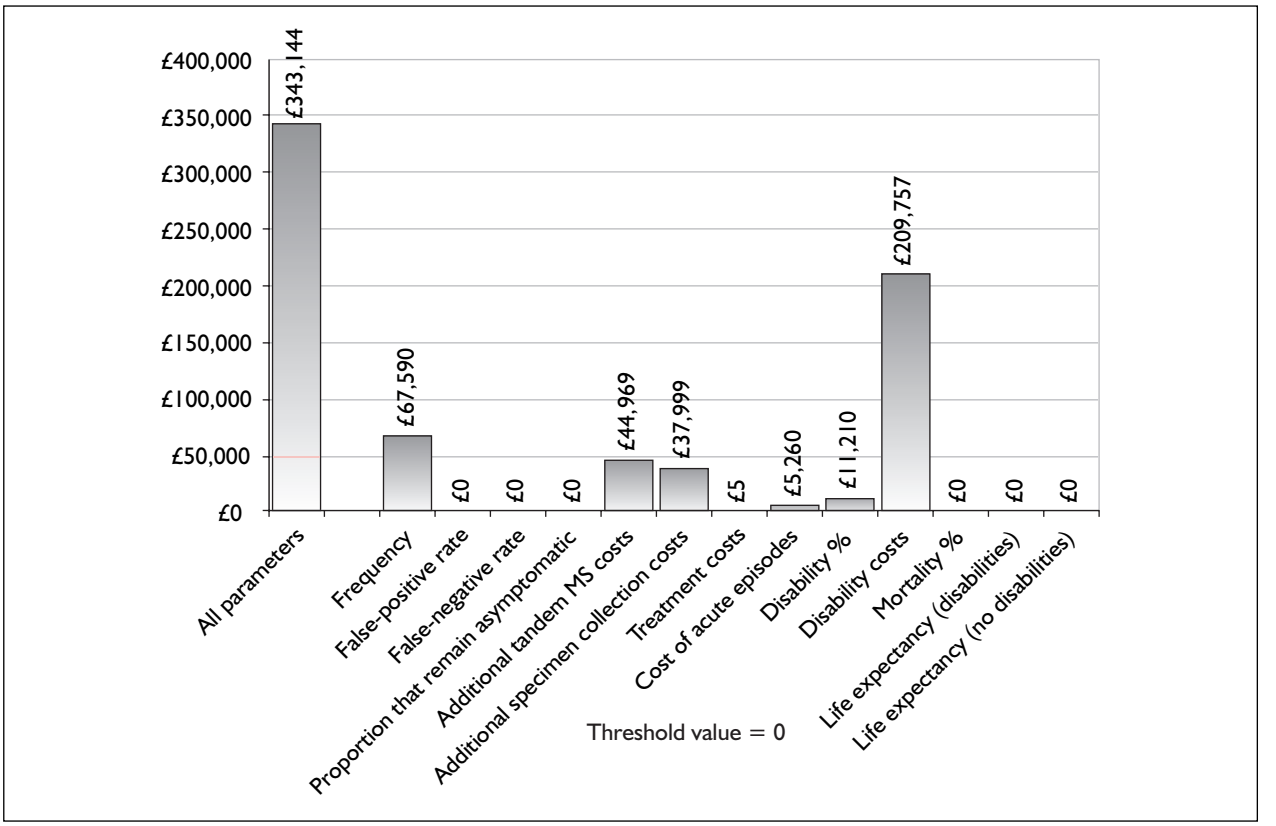

FIGURE 12 *EVPI for MCAD deficiency model parameters (threshold = 0)*

net benefit given perfect information on the parameter of interest minus the average net benefit given existing information.

The EVPI results indicate that the most significant model parameters to affect the expected net benefit in the decision model for MCAD deficiency are the frequency (incidence) of the condition and the future costs related to disabilities. Within the total UK context, establishing the true underlying incidence of MCAD deficiency detectable through screening could only really be determined accurately through a prospective evaluation of screening practice. This would have to include a wide range of participating centres because of the known regional variations in incidence. The EVPI analysis also shows that another major area of uncertainty to impact on model results is the extent to which specific disorders lead to significant (avoidable) disabilities and the future resource costs that these impose on the healthcare, social care and educational sectors. It is important to note that uncertainty around model estimates of the proportion of symptomatic cases that develop disabilities is not significant for the EVPI analysis. This should not be interpreted to mean that the proportion of symptomatic cases who develop disabilities is not a major driver of cost-effectiveness of screening, only that the level of uncertainty around the estimates of this proportion used in the model (10–15%) is not sufficiently wide enough to have a significant impact on the results of the decision model (i.e. on the decision whether to accept the screening or no-screening options evaluated). Similar issues arise in respect of the proportion of cases who remain asymptomatic and mortality rates in symptomatic cases for MCAD deficiency. If levels of uncertainty on these parameters were significantly greater than has been represented within the models, then this would require revised estimates to be included. However, as most of the information has been based on published research, it is felt that the economic models characterise the current levels of uncertainty around these issues sufficiently well.

The implicit assumption of adopting a threshold value of zero, however, should be clearly understood before any decision to fund additional research on the basis of this analysis is made. A zero threshold value implies that society is not willing to pay more for any extra health benefits that may accrue from a new technology (in effect, the value of an additional life-year saved through

MCAD deficiency screening is set at zero). This does not seem a realistic assumption.

Discussion

The results of the economic modelling presented here provide useful evidence on the potential cost-effectiveness of using tandem MS for screening of neonatal disorders of metabolism. The following discussion is organised into four main sections: costs, benefits, service delivery issues, and the implication of the results presented here for future research.

Costs

Substituting tandem MS for existing technologies used in the UK neonatal screening programme for PKU is not cost-effective. Compared with the existing profile of modalities used for PKU screening, tandem MS is more expensive for no additional gain in health benefits.

Nevertheless, the distinctive feature of tandem MS as a screening technology is that the marginal costs of screening for a number of additional disorders in the laboratory are relatively small, certainly when compared with some of the potential health gains available. There are, however, other costs associated with a decision to screen for additional disorders. The incremental costs related to a specific inborn metabolic condition need to be carefully evaluated in terms of potential resource saving attributable to presymptomatic detection of affected children and any associated health gains obtained.

This economic evaluation has specifically sought to examine whether the adoption of tandem MS could be cost-effective for an extension of the current neonatal screening programme. With this in mind, it is appropriate that the full costs of a tandem MS system are included as a cost. Nevertheless, this overlooks the fact that many tandem MS instruments used in a neonatal screening programme will also be used for other purposes besides the initial screening process. The uses include continued diagnostic work and post-mortem examinations for other metabolic conditions not screened for. Tandem MS is also used for monitoring and controlling the diets of patients with PKU.

It is also worth repeating that using estimates based on the distribution of costs from the previous HTA report implicitly assumes that the costs associated with the 'existing technologies' option have the same relative combination of screening modalities and operating efficiencies as obtained in 1996. To the extent that practice has shifted to more expensive capital-intensive methods, then the probabilistic estimates used within the model may understate the cost of the existing technologies option. With more capital-intensive methods, however, services may be able to take advantage of economies of scale and therefore the estimates included in the model may overstate the cost of current technologies combined.

Current practice is unlikely to be precisely the same as it was when the earlier HTA review approached laboratories for details of operating costs. At that time significant variations in operating costs were obtained, even within the same screening technology used. This variation has been incorporated within the economic model using a probabilistic estimate that sampled across this range of cost estimates. How substantially different current practice across all laboratories is today – in terms of methods used, as well as scale, operational efficiency and costs – could only be ascertained with certainty through a detailed updated review of all existing participating centres.

However, given the very high probability of the cost-effectiveness of extending screening to include MCAD deficiency there would need to be quite a significant difference in the cost of providing existing technologies compared with the estimates obtained in the model to alter the overall results dramatically.

The model was constructed so that (future) disability costs were applied only to those who would become symptomatic and survive early death (i.e. the expected number of early deaths was subtracted from the symptomatic caseload before estimates of disability were applied). However, in some disorders, patients who die in early infancy will have already incurred some significant additional costs before the event; partly as a direct result of acute hospitalisation, but also because many of those who die may have been affected by severe neurological damage disabilities. To the extent that this may occur, the model will underestimate the cost savings attributable to screening.

Indirect costs have not been considered in the present model. Nevertheless, future indirect costs to patients affected by disabilities caused by metabolic disorders not recognised early, and indirect costs to their families and informal carers, would be significant.

Benefits

Health benefits within the economic model described here were expressed in terms of gains in life-years only. No account was taken of the quality of those additional life-years. This could be seen as a significant limitation if the additional years gained from screening were of relatively poor quality. However, there is evidence to suggest that for the highest priority condition examined here, MCAD deficiency, the quality of life of individuals who are diagnosed and treated is high. The absence of QALY weights for MCAD deficiency almost certainly underestimates the health benefits of screening because no account is taken of reduced quality of life for those affected by moderate to severe disabilities, most of which should be avoided through the early detection and treatment afforded by screening.

The model construction assumes that mortality in MCAD deficiency cases is due entirely to symptomatic presentation. However, there is evidence of occasional deaths in cases already diagnosed. Nevertheless, as reported in the clinical review (Chapter 6; MCAD deficiency, 'Discussion', p. 55), various studies of infants with MCAD deficiency have found that, following early diagnosis and initiation of effective treatment, no deaths have occurred[20,23,28,63] and there have been no cases of appreciable cognitive impairment or neurological damage.[28,63] Therefore, this assumption appears valid.

It is acknowledged that being symptomatic is not necessarily synonymous with being diagnosed and that a proportion of cases who show signs of inborn metabolic disorders (including MCAD deficiency) may go unrecognised. A proportion of these cases may subsequently develop disabilities and complications. The evidence on the extent to which this may occur in practice is limited. However, for MCAD deficiency, this problem will only occur to any significant extent in the absence of screening. As such cases would incur avoidable future costs, the current economic model will further tend to underestimate the relative cost-effectiveness of using tandem MS for neonatal screening of PKU and MCAD deficiency combined.

It is also recognised that restricting the measurement of benefits in terms of life-years gained omits important psychological impacts associated with screening. However, a review of the available evidence in a previous HTA review[1] suggests that, on balance, the psychological benefits outweigh the costs.

Service delivery and organisation

Laboratory services

Adopting tandem MS technology would have some major implications for the organisation of neonatal screening laboratories in the UK. The evidence presented here clearly demonstrates that there are significant scale economies using tandem MS technology. However, what constitutes an appropriate volume of specimens to be processed at an individual screening laboratory is not only constrained by issues of operational feasibility. Given that the total number of samples that require analysis in a given year is limited, the number of participating laboratories would affect the volume of specimens processed at each laboratory. In these circumstances there would be a real trade-off between technical efficiency (economies of scale and scope) and the organisation of existing and future neonatal screening services. This trade-off was recognised in the previous HTA review that examined the potential use of tandem MS. That report concluded that:

> "Laboratories undertaking tandem MS screening should have adequate workload to justify the capital investment and staffing required. A minimum of 50,000 samples per year is suggested."[1]

With approximately 700,000 live births in the UK at present, this would only require some 12 or 13 neonatal screening providers operating at around 50,000–60,000 samples per year.

In setting out the future direction of its Pathology Modernisation programme, the Department of Health drew attention to recent reviews of pathology services that "identified considerable scope for strengthening and streamlining the pathology service, subject to the availability of capital for investment".[101] As part of the modernisation fund, specific resources have been provided to "modernise pathology services and encourage larger systems of clinical pathology that will have the capacity to deal with rising demand and technological developments".[101] Therefore, the process of reconfiguring UK pathology services is already underway.

Screening for additional disorders, which require early collection of samples for best results (e.g. MCAD deficiency), may require additional changes in the way laboratories are organised in order to facilitate a rapid response. Any further extension of screening would also require neonatal screening laboratories to work closely with clinical teams.

Paediatric services
There would also be significant implications for the resourcing and organisation of paediatric services, particularly the availability of specialists in the management of relatively rare metabolic disorders. There would need to be a well-structured clinical network to assess, diagnose and treat all babies identified in the screening programme. Families of affected babies would need to receive appropriate counselling follow-up.

Screening for additional disorders would change the workload patterns for the small number of specialists in paediatric metabolic medicine. Currently there is a shortage of these in training and there are few clinical nurse specialists.

Overall, however, screening could provide some compensating resource effects on paediatric services more generally through the avoidance of expensive hospitalisations that can occur with symptomatic cases presenting in acute metabolic crisis.

Community services
There would also be important service delivery issues for community services if neonatal screening were extended to other conditions. Midwives and health visitors would require some extra training to be able to provide appropriate advice and explanations to parents.

Additional information about screening for other disorders would also need to be provided at antenatal contact to prepare parents for screening interventions in the neonatal period.

Again, the working practices of midwives and health visitors might need to be examined to ensure that the collection of samples is undertaken at the most appropriate time and that possible delays are minimised.

Research
Particular issues to emerge from the EVPI analysis are that the major drivers of the cost-effectiveness of using tandem MS for MCAD deficiency are incidence and future health and welfare costs related to disabilities and complications. However, the results of the economic modelling of neonatal screening using tandem MS for both PKU and MCAD deficiency combined appear to be robust. This explains why the overall value of additional research information, suggested by the EVPI analysis, is low. Adoption of tandem MS technology for both PKU and MCAD deficiency screening has a very high probability of cost-effectiveness, even at low threshold values for additional life-years gained.

Although the economic modelling of GAI and homocystinuria (CβS) indicates a high probability of cost-effectiveness at relatively low threshold values, there is less confidence in the results obtained. For GAI there remain some significant uncertainties. Although GAI is a rare condition, there is still a lack of published data on the underlying incidence in the UK, and limited published data on the efficacy of the treatment regimens to prevent disabilities and complications after diagnosis; further additional data are needed on the sensitivity and specificity of detection using tandem MS. For homocystinuria (CβS), although there is more recent data on the efficacy of treatment, it is also a rare disease and published evidence suggests that it may not be reliably detected using tandem MS. For both conditions, additional economic modelling work should be undertaken, potentially drawing on the clinical knowledge and expertise of those currently working in the field in the UK, to resolve these uncertainties.

For many of the other conditions potentially detectable by tandem MS, there remains an absence of good economic data on the potential future treatment, acute care and disability costs associated with presymptomatic and symptomatic detection. Additional research on these conditions should focus on providing these data.

Information on the HRQoL of affected patients, particularly when stable and during occasional episodes of decompensation, would allow for a more sensitive measurement of the potential future health outcomes for individuals treated. Similar information on quality-of-life impacts on families is also lacking.

Research using a health utilities framework could also provide information for the quantification of many of the other significant positive and negative non-resource effects caused, or avoided, by screening. However, the priority would be to undertake this work in respect of all the other conditions where additional evaluation work is required around the relative cost-effectiveness of including them in a neonatal screening programme (i.e. excluding MCAD deficiency).

Chapter 8

Implications for other parties

Pollitt and colleagues[1] comprehensively discussed the psychological implications of diagnosis through neonatal screening. These authors concluded that, on balance, the psychological benefits of screening outweigh the costs. Some of these issues, affecting parents, are discussed below.

Early detection of a genetic disorder in a newborn infant, through neonatal screening, is likely to be a great shock and an emotionally traumatic experience for parents, for whom there may have been no warning signs. Conversely, early diagnosis of a problem may be helpful for parents, allowing them to prepare emotionally and practically for the development of the disease in the child.[1]

False diagnoses involved in neonatal screening can also create anxiety and psychological problems for parents. In addition to the emotionally traumatic experience of receiving a false diagnosis of a genetic disorder, some parents may have lingering concerns that the initial diagnosis may have been correct, even after the diagnosis has been overturned by repeat testing.[1] Although neonatal screening tests are designed to detect infants with metabolic illnesses, certain tests may also identify carriers or individuals with variants who may be clinically asymptomatic.[102] For example, an unknown proportion of biochemically affected individuals with MCAD deficiency will never experience any ill effects and there is no way of predicting which individuals these will be.[1] Potential consequences of diagnosis for this asymptomatic group include anxiety about the risk of hypoglycaemia during early childhood and adverse effects of clinically unwarranted treatment.[103] In a potential newborn screening programme, there may be a tendency to regard asymptomatic cases as false positives; however, such individuals are at as much risk as the symptomatic cases but lucky not to have encountered a sufficient metabolic stress to precipitate a crisis. Therefore, all babies with MCAD deficiency detected by newborn screening are at risk and treatment is warranted in all.[104] Pollitt reported that reasonable parental anxiety about the possibility of hypoglycaemia during early childhood is an appropriate, indeed desirable, consequence of a screening diagnosis.[104] In general, parents are supportive of screening despite any stress it may cause.[1,104]

Other consequences of diagnosis for the asymptomatic group include genetic information for parents. Such information is important for the family because of the identification of carrier couples. This may have an impact on future pregnancies and other family members. Paediatricians need to be aware of this aspect of newborn screening and provide appropriate counselling or referral of such families.[102]

Pollitt and colleagues[1] reported that very few parents understand the concept of neonatal screening in terms of diseases tested for and the screening process involved for disease detection. Parental anxiety and dissatisfaction with the screening process have been found to be greater in those who feel they have not received adequate information. The provision of good, clear information about the processes involved in neonatal screening is vital and may ease the unnecessary anxiety that some parents feel. Therefore, information regarding why a blood sample is being taken, what disorders are being tested for and what the results will mean for the child should be provided to parents and the delay in reporting a correct diagnosis, both positive and negative, should be kept to a minimum. Information about false-positive (potentially the cause of much distress) and false-negative test results should also be provided to reduce the fears and anxiety that these can create for parents.[1]

In the UK, informed consent may not always be obtained in relation to neonatal screening and parents may not receive enough information to make a truly informed choice about whether to have their child screened for a genetic disorder. Obtaining informed consent from parents through the provision of comprehensive information about screening may reduce parental anxiety.[1]

Chapter 9

Factors relevant to the NHS

Pollitt and colleagues[1] and Seymour and colleagues[2] comprehensively discussed many of the issues and organisational arrangements for the (expansion of the) current neonatal screening programme in the UK. Some of these are discussed below.

Specimen collection

In the UK, the neonatal screening programme collects heel-prick blood samples from babies at 6–14 days of age,[105] to maximise the sensitivity for the PKU screen. The specimen is collected either as dried blood spots or as liquid blood depending on the local practice. This time-lapse may be detrimental for disorders that present acutely in infancy and there may be a case eventually to move to an earlier collection. In addition, non-UK newborn screening programmes using tandem MS generally collect the blood sample much earlier than current practice in the UK and return the result within 2 days. This would mean 7-day working in laboratories, which would need to work much more closely with clinical teams (Leonard J: personal communication; 2002).

Ades and co-workers[106] found that in the North Thames region, the timeliness of neonatal screening showed unacceptable variation between districts and delays in dispatch of specimens to the laboratory. These authors concluded that same day, first-class posting should be introduced and samples collected between 4 and 8 days. Cost was not a major consideration as most Guthrie cards were already sent first class. A move to sample collection between 4 and 8 days would fit well into the routine midwife health checks, which take place every day for the first week of life.[106]

The evidence obtained in this review suggests that the collection of newborn blood spots in the UK at slightly earlier times than the current 6–14 days might facilitate earlier and clinically more effective therapy. However, owing to changes in concentrations of various metabolites over time, the earlier collection and reporting of results may influence the sensitivity and specificity of the screening process for other conditions. Any changes would underline the need for a better infrastructure for clinical follow-up, management and high-quality clinical services for identified cases and their families.[2,107] As a separate measure, Pollitt and colleagues[1] reported that selective screening for all sick neonates should also be performed with minimal delay as this approach may lead to a cost-saving in intensive care bed occupancy.

Coverage and record-keeping

There is a lack of accurate data on the coverage of neonatal screening in the UK. Ades and colleagues[10] found that the coverage of neonatal screening in the UK was higher than recent reports suggest, but it was difficult to distinguish failure to screen from failure to record screening. The new system for issuing NHS numbers at birth (printed on Guthrie cards), which was implemented in October 2002, provides a basis for a reliable information system of neonatal screening on a national scale. This will require that maternity, neonatal laboratory and child health computers are all linked into a national network to ensure that testing is carried out and results are recorded regardless of where parents are living, where the sample is taken and where the tests are carried out.[10]

Provision of information and ethics

Issues relating to the provision of information to parents and informed consent were raised in Chapter 8. In brief, the provision of good, clear information about the processes involved in neonatal screening is vital and may ease the unnecessary anxiety that some parents feel. For a screening test to be fully effective its sensitivity should ideally approach 100%; however, owing to the element of human error and the potential for biological variability, no screening test sensitivity can truly achieve 100% over long periods.[108] Therefore, screening programmes are always likely to fall short of detecting all cases of a screened condition and will have a limited sensitivity for a number of diseases to be screened. To avoid litigation over missed cases, recipients of screening or their carers need to be made clearly aware that screening has the potential to save lives or improve quality of life through early diagnosis;

however, it cannot guarantee to detect all cases of a screened condition. Detailed information should also be made available to professionals regarding the sensitivity and specificity of screening for some disorders.

The principle of informed consent presents particular difficulties for neonatal screening. Screening in the UK has always been voluntary; however, few places in the UK provide adequate information to parents to make a truly informed choice and few require informed parental consent for newborn screening. However, parents are entitled to be informed and to give consent before samples are taken.[2]

Clinical referral and treatment

Many of the conditions detectable by tandem MS are severe and would require a streamlined approach to referral and treatment. There would need to be a well-structured clinical network to assess all babies identified in the screening programme so that correct diagnosis and treatment could be established. The family would need to be counselled satisfactorily with appropriate follow-up. This would change the pattern of the workload for the small number of paediatricians in paediatric metabolic medicine, who are currently thinly spread across the UK. There is a shortage of those in training and there are few clinical nurse specialists (Leonard J: personal communication; 2002).

Pollitt and colleagues[1] reported that "as many of the patients will be asymptomatic, emergency referral to a distant specialist hospital may be inappropriate, but the local paediatrician must be provided with clear advice about warning signs and management before specialist assessments. For some of the 'new' acutely presenting disorders, the mainstay of treatment is the avoidance of catabolic stress and in more remote areas, it will be necessary to educate both the family practitioner and parents in emergency measures."

Laboratory services

Laboratories undertaking tandem MS screening would need to have an adequate workload to justify the capital investment and staffing required. The economic analysis suggests that an appropriate (average) operating volume for each tandem MS system would be in the range of 50,000–60,000 samples per year. This would have significant implications on the number of laboratories providing a neonatal screening service in the UK. In addition, these laboratories would need staff and scientists capable of running and maintaining the instrument, interpreting the results and undertaking or organising the necessary confirmatory investigations.[1]

False diagnoses involved in neonatal screening can create anxiety and psychological problems for parents. There is a need to establish approaches and cut-off decisions for identifying a reasonably firm presumptive positive test result without the need for repeat testing or referral of the patients, and the delay in reporting a correct diagnosis, either positive or negative, should be kept to a minimum. Automated flagging and processing in tandem MS would speed up the neonatal screening process and reduce labour costs.

In many areas of the UK, there is relatively poor liaison between the screening laboratories, midwives and healthcare personnel. Seymour and colleagues[2] reported that there was a clear need for a better infrastructure (clinical consultants, scientists in paediatric metabolism, metabolic investigation laboratories, paediatric dieticians and clinical genetic services) for the confirmation of abnormal results, other specific diagnostic tests, notification and continued care (including parental counselling) and management of patients with inborn errors of metabolism. The expansion of newborn screening programmes in the UK would depend on this infrastructure.

Other factors

Hutchesson and colleagues[30] reported that documentation of both ethnicity and the family tree in the records of all patients with inborn errors of metabolism should be encouraged. This information is also important for those involved in the care of patients and in the genetic counselling of their relatives and valuable in monitoring patterns of diseases in the community.

Chapter 10
General discussion

Clinical and cost-effectiveness review

This HTA report included a systematic review of the literature and an analysis of the available economic evidence for neonatal screening of inborn errors of metabolism using tandem MS. The results of the clinical effectiveness and cost-effectiveness analyses support the introduction of tandem MS into a UK neonatal screening programme for PKU and MCAD deficiency combined. However, at present the evidence does not support the extension of the UK neonatal screening programme to include all disorders detectable by tandem MS.

Clinical evidence

The evidence on the clinical effectiveness of neonatal screening for inborn errors of metabolism using tandem MS comes primarily from observational data of large-scale prospective newborn screening programmes, from several centres outside the UK, namely, Australia, Germany and the USA. RCTs of screening for rare disorders are difficult because of the need for a prolonged follow-up period[105] and the enormous numbers that would be needed for adequate power.[1,4,20]

In general, newborn screening of dried blood spots for the amino acid and acylcarnitine group of disorders using tandem MS has been shown to be rapid, highly sensitive (90–100%) and highly specific (99–100%). However, direct comparison of the performance of tandem MS between studies is limited owing to the variation in the age of sampling between studies and the use of various thresholds (based on percentiles, concentrations of compounds and ratios) to identify inborn errors of metabolism. The diversity in the preference of metabolite together with the cut-off limits used to define a positive result also affects the yield of false-positive and false-negative results. These major differences between studies make it difficult to draw firm conclusions from this evidence.

The sensitivity and specificity of neonatal screening for individual inborn errors of metabolism using tandem MS are limited, except for MCAD deficiency. Several large-scale prospective newborn screening programmes from Australia, Germany and the USA and retrospective data from the UK (because of the moratorium on extending screening programmes in the UK, only retrospective studies are available) have, in general, illustrated high sensitivities (100%) and high specificities (100%) of neonatal screening for MCAD using tandem MS. However, direct comparison of the diagnostic performance and outcome of tandem MS between studies is limited by the use of various analytes and thresholds for detecting disease status and different criteria for the diagnosis of MCAD deficiency.

The UK newborn screening programme for PKU is well established and there is universal agreement that neonatal screening for PKU is justified. The average UK incidence of PKU (classical and atypical combined) is 11.0 cases per 100,000 live births. Early dietary interventions, including dietary treatment before or during pregnancy, are effective in reducing the severity of developmental delay. Severe retardation due to PKU has all but disappeared in the screened population.[1]

Of the many other amino acid and acylcarnitine disorders detectable by tandem MS, the evidence supports the introduction of MCAD deficiency in a tandem MS-based neonatal screening programme.

MCAD deficiency, a disorder of fatty acid metabolism (the expected UK birth prevalence/incidence ranges from 4.0 to 9.9 cases per 100,000 live births) is associated with increased morbidity and mortality. Treatment for MCAD deficiency is relatively simple with dietary management, thus preventing possible early death and neurological disability. After (early) diagnosis, current management makes death rare and improves outcome.[20,23,60] However, the presentation of MCAD deficiency varies widely (the vast majority of patients present with metabolic crisis during the first years of life when metabolically challenged by fasting and/or viral illness.[23] Although a few children are clinically diagnosed with MCAD deficiency in the neonatal period, episodes usually occur between 3 months and 3 years of age, with an average age of 12 months),[1,109] with some individuals not presenting until they are adults and an unknown number

remaining undiagnosed or asymptomatic.[107] Potential consequences of diagnosis for this asymptomatic group include anxiety about the risk of hypoglycaemia during early childhood and adverse effects of clinically unwarranted treatment.[103] Pollitt and colleagues[1] reported that all infants with MCAD deficiency detected by newborn screening must be treated as being at risk and treatment is warranted in all. The need to monitor carefully intercurrent illnesses for the first few years of life will cause parental anxiety; however, there is also the possibility that "a good counselling and support service with experienced clinicians can reduce this" (Leonard J: personal communication; 2002). Sensitivity and specificity of neonatal screening for MCAD deficiency using tandem MS have been shown to be high for babies less than 72 hours of age,[7,20,22,23] but Clayton and co-workers[28] conclude that if screening is undertaken at 7–10 days of age the number of false-positive and false-negative results should be negligible. For other inborn errors of metabolism that can be detected by tandem MS, robust clinical evidence is limited.

In most of the large-scale prospective newborn screening programmes using tandem MS, internationally, sampling for selected amino acids and acylcarnitines was usually performed within 72 hours after birth, whereas in the UK, collection of blood samples is usually done at about 6–14 days of age. The age at which screening is undertaken will affect the sensitivity and specificity of the screening process as concentrations of metabolites change over time and this time-lapse may be detrimental for disorders that present acutely in infancy. The evidence obtained in this review suggests that the collection of newborn blood spots in the UK at slightly earlier times than the current 6–14 days might facilitate earlier and clinically more effective therapy. However, the earlier collection and reporting of results may influence test performance for other conditions and would underline the need for a better infrastructure for clinical follow-up, management and high-quality clinical services for identified individuals and their families.[2,107] More research would be needed in this area if tandem MS-based neonatal screening were introduced for the identification of further disorders.

The clinical evidence found that there have been no controlled trials of treatment for the vast majority of inherited metabolic diseases, but some type III studies demonstrated efficacy of treatment. The general standard of reporting in these studies was poor, with inadequate details about research methods and relevant clinical details. This makes it difficult to assess the validity and relevance of findings. In addition, it must be recognised that the some of the information in this report, even for the more common conditions, may not reflect current practice in the UK, and there is a paucity of robust data on long-term outcomes. The clinical evidence for many of the inherited metabolic disorders suggests that there are uncertainties regarding the effectiveness of treatments, the symptom-free interval for some of these disorders is very short, the disorders are more heterogeneous than MCAD deficiency and the natural history is less well established. For many of the individual inborn errors of metabolism, although these are relatively rare, there is a lack of data regarding the underlying incidence in the UK and more work is needed on the sensitivity and specificity of detection using tandem MS.

Although newborn screening has the potential to save lives or improve quality of life through early diagnosis, it cannot guarantee to detect all cases of a screened condition. The delay in reporting a correct diagnosis, either positive or negative, should be kept to a minimum and the provision of good, clear information about the processes involved in neonatal screening, explanation of results (including information about false-positive and false-negative test results), genetic counselling, recessive diseases and consanguinity,[30] is vital and may ease the unnecessary fear and anxiety that some parents feel.[1]

Any expansion of newborn screening programmes in the UK would depend on a better infrastructure (clinical consultants, scientists in paediatric metabolism, metabolic investigation laboratories, paediatric dieticians and clinical genetic services) for the confirmation of abnormal results, other specific diagnostic tests, notification and continued care (including parental counselling) and management of patients with inborn errors of metabolism identified through newborn screening.[2]

Economic analysis

The results of the economic modelling presented in this report provide evidence that the use of tandem MS for the combined screening of PKU and MCAD deficiency would be cost-effective. This conclusion is based on probabilistic modelling of data derived from a number of sources. It is acknowledged that there remains uncertainty around some of the parameter values used in the model. However, one of the major advantages of the current approach is that the impact of this

uncertainty on the results can be included and quantified.

One important consideration for any new screening programme is the extent to which the screening technology generates false-positive results. Such false positives incur additional resource costs and may impose significant psychological disbenefits on those affected (especially families). In relation to the use of tandem MS screening of PKU and MCAD deficiency, the overall evidence suggests that the technology is highly sensitive and specific. Nevertheless, there is still some uncertainty around this parameter and so the economic model incorporated estimates based on including a range of values of the number of false positives generated within any cohort of 100,000 neonates screened. The additional resource consequences of these false-positive results are included in the model. In terms of resource costs, the VOI analysis indicates that the potential number of false positives generated by tandem MS in a combined PKU and MCAD screening programme does not have a significant impact on overall cost-effectiveness.

No evidence (i.e. data) was included to measure the potential psychological disbenefits associated with false-positive results. However, any negative quality of life effects due to false positives must be set against the significant (long-term) quality of life effects of adverse outcomes preventable through presymptomatic detection provided by screening. Given the conclusions of the review of psychological evidence in the Pollit report[1], the balance of quality of life effects would favour screening for PKU and MCAD deficiency.

It is recognised that specific aspects of a number of the treatment strategies used within the economic model may not reflect current practice in some respects. For example, there may be some differences in current practice regarding the number of years for which treatment may be given and over the use of L-carnitine supplementation for MCAD deficiency. Once again, however, these uncertainties were characterised in the model using range estimates of treatment costs. For MCAD deficiency, the range was based on an update of the base-case cost used by Pollit and colleagues[1] and four times this figure. The VOI analysis suggests that this wide variation in the potential costs incurred for treatment does not have a major impact on the overall results obtained.

The most important model parameter to affect the results is the magnitude of future healthcosts and social care costs that arise from disabilities and complications, even though these may only affect a small number of individuals.

Most of this uncertainty reflects assumptions about the volume and valuation of the future costs incurred, and some of this is due to the underlying variation in severity of disabilities that may arise from symptomatic presentation. The degree of uncertainty around these variables is reflected in the wide variation in future costs potentially avoided through screening (for each cohort of 100,000 neonates). The VOI analysis clearly indicates that this has the largest single impact on the economic case for using tandem MS in a combined PKU and MCAD deficiency screening programme. Despite this uncertainty, the model produces a high probability of cost-effectiveness for PKU and MCAD deficiency screening combined at relatively low threshold values.

It is important to reiterate that this result is based on an assumed utilisation rate of between 50,000 and 60,000 samples per tandem MS system. At volumes of 20,000 samples per system the probability that using tandem MS would be cost saving is zero. However, it would be inefficient to use such capital-intensive methods at volumes significantly below capacity. At operating volumes of between 50,000 and 60,000 samples per year the model suggests a modest incremental cost saving from the use of tandem MS for both PKU and MCAD deficiency. This represents a realistic operating range for a single tandem MS system. Adoption of tandem MS technology for the UK screening programme, therefore, would have significant implications for the future configuration of neonatal screening laboratories.

Need for further research

- The economic evidence concerning the use of tandem MS for PKU and MCAD deficiency combined is very favourable. The key assumptions underlying this analysis that may constitute areas for further research are the future disability costs associated with symptomatic cases, and the underlying population incidence of the condition in England and Wales. However, the VOI analysis undertaken as part of the economic modelling of cost-effectiveness suggests that the overall value of obtaining these additional data is not high.
- Tandem MS has the potential for simultaneous multidisease screening using a single analytical

technique. However, it is difficult to draw firm conclusions on extending the UK neonatal screening programme to all disorders detectable by tandem MS. More research is needed to establish the sensitivity and specificity of neonatal screening using tandem MS for other individual inborn errors of metabolism in the UK.

- Although the economic modelling of GAI and homocystinuria (CβS) indicates a high probability of cost-effectiveness, some significant uncertainties remain. For both conditions, additional economic modelling work should be undertaken, potentially drawing on the clinical knowledge and expertise of those currently working in the field in the UK, to resolve these uncertainties.
- Long-term studies are needed to evaluate whether outcomes are improved as a result of tandem MS screening and early diagnosis. To conduct adequate long-term follow-up, programmes need to be established or patient tracking systems improved. Ideally, data management for such a system would include registries to which treatment centres continually provide updated information, treatment compliance and outcomes.[6] An extension of the existing register for PKU should be established, to gather further data about the underlying incidence and natural history of other inborn errors of metabolism, which may in the future be the subject of a screening programme.
- Further research is needed to ascertain the natural history of the conditions, and the economic impact, for the other metabolic disorders. The primary focus of this research should be on the long-term effectiveness of treatment strategies on adverse outcomes (disabilities and impairments) under conventional management and the potential impact of early diagnosis using tandem MS.
- Research into the measurement of a range of intangible (psychological) benefits and disbenefits associated with screening, using either a health utilities or a willingness to pay framework, should be considered. This would provide a more appropriate basis on which to compare and quantify the net impact of the non-resource consequences of neonatal screening for other inherited metabolic disorders.

Chapter 11
Conclusion

This systematic review evaluated the clinical effectiveness and cost-effectiveness of neonatal screening for inborn errors of metabolism using tandem MS. The evidence supports the introduction of tandem MS into a UK neonatal screening programme for PKU and MCAD deficiency combined.

New technological approaches for automated processing, coupled with computer-assisted software, would allow the analysis of hundreds of samples on a daily basis and minimise labour costs. Tandem MS has the potential for simultaneous multidisease screening using a single analytical technique. However, it is difficult to draw firm conclusions on extending the UK neonatal screening programme to all disorders detectable by tandem MS. Although the marginal cost of extending the programme to include other conditions may be relatively small, the application of this new technology to PKU and MCAD deficiency screening does not imply the wholesale inclusion of all disorders detectable by tandem MS. Robust evidence on the underlying incidence and outcomes for many of the disorders was lacking, particularly differences in long-term outcomes that could be attributed to therapies initiated as a consequence of presymptomatic detection using tandem MS.

Acknowledgements

This review is an update of two existing systematic reviews of neonatal screening for inborn errors of metabolism:

Pollitt RJ, Green A, McCabe CJ, Booth A, Cooper NJ, Leonard JV, *et al*. Neonatal screening for inborn errors of metabolism: cost, yield and outcome. *Health Technol Assess* 1997;**1**(7).

Seymour CA, Thomason MJ, Chalmers RA, Addison GM, Bain MD, Cockburn F, *et al*. Newborn screening for inborn errors of metabolism: a systematic review. *Health Technol Assess* 1997;**1**(11).

The authors wish to thank Dr Anne Green, Consultant Clinical Biochemist, Birmingham Children's Hospital, Professor Rodney Pollitt, Neonatal Screening Laboratory, Sheffield Children's Hospital, Dr Mark Sharrard, Sheffield Children's Hospital, Dr Darren Shickle, Clinical Senior Lecturer in Public Health Medicine, ScHARR, University of Sheffield, and Professor Stuart Tanner, Institute of Child Health, Sheffield Children's Hospital, UK, for their helpful advice. The authors also wish to thank Andrea Shippam for her help in preparing and formatting the report.

This report was commissioned by the NHS R&D HTA programme. The views expressed in this report are those of the authors and not necessarily those of the NHS R&D HTA programme. All responsibility for the content of the report remains with the authors.

Contributions of authors

Abdullah Pandor (Research Associate, Rapid Reviews Group, ScHARR) was responsible for the review of clinical effectiveness, Joe Eastham (Senior Research Officer, Nuffield Institute for Health) was responsible for the economic analyses and Catherine Beverley (Systematic Reviews Information Officer, ScHARR) undertook the electronic literature searches. Jim Chilcott (Technical Director, Rapid Reviews Group, ScHARR) provided help in developing the Expected Value of Perfect Information (EVPI) model and Suzy Paisley (Managing Director, Rapid Reviews Group, ScHARR) provided guidance in the production of the report. All authors assisted in drafting the report.

About ScHARR

The School of Health and Related Research (ScHARR) is one of the four Schools that comprise the Faculty of Medicine at the University of Sheffield. ScHARR brings together a wide range of medical- and health-related disciplines including public health, general practice, mental health, epidemiology, health economics, management sciences, medical statistics, operational research and information science. It includes the Sheffield unit of the Trent Institute for Health Services Research, which is funded by NHS R&D to facilitate high-quality health services research and capacity development.

The ScHARR Technology Assessment Group (ScHARR-TAG) synthesises research on the effectiveness and cost-effectiveness of healthcare interventions for the NHS R&D Health Technology Assessment Programme on behalf of a range of policy makers, including the National Institute of Clinical Excellence. ScHARR-TAG is part of a wider collaboration of six units from other regions. The other units are: Southampton Health Technology Assessment Centre (SHTAC), University of Southampton; Aberdeen Health Technology Assessment Group (Aberdeen HTA Group), University of Aberdeen; Liverpool Reviews & Implementation Group (LRiG), University of Liverpool; Peninsular Technology Assessment Group (PenTAG), University of Exeter; NHS Centre for Reviews and Dissemination, University of York; and West Midlands Health Technology Assessment Collaboration (WMHTAC), University of Birmingham.

References

1. Pollitt RJ, Green A, McCabe CJ, Booth A, Cooper NJ, Leonard JV, *et al*. Neonatal screening for inborn errors of metabolism: cost, yield and outcome. *Health Technol Assess* 1997;**1**(7).
2. Seymour CA, Thomason MJ, Chalmers RA, Addison GM, Bain MD, Cockburn F, *et al*. Newborn screening for inborn errors of metabolism: a systematic review. *Health Technol Assess* 1997;**1**(11).
3. National Screening Committee. First report of the National Screening Committee. London: NSC; 1998. (http://www.doh.gov.uk/nsc/library/lib_ind.htm).
4. Wilcken B, Wiley V. Tandem mass spectrometry in the New South Wales newborn screening program. *MMWR Morb Mortal Wkly Rep* 2001;**50** (RR-3): 33.
5. Wilcken B, Wiley V, Carpenter K. Two years of routine newborn screening by tandem mass spectrometry (MSMS) in New South Wales, Australia. *J Inherit Metab Dis* 2000;**23** (Suppl 1):4.
6. Centers for Disease Control and Prevention. Using tandem mass spectrometry for metabolic disease screening among newborns: a report of a work group. *MMWR Morb Mortal Wkly Rep* 2001;**50** (RR-3):1–36.
7. Zytkovicz TH, Fitzgerald EF, Marsden D, Larson CA, Shih VE, Johnson DM, *et al*. Tandem mass spectrometric analysis for amino, organic, and fatty acid disorders in newborn dried blood spots: a two-year summary from the New England Newborn Screening Program. *Clin Chem* 2001;**47**:1945–55.
8. Muenzer J, Frazier DM, McCandless SE. Incidence and follow-up evaluation of metabolic disorders detected by newborn screening in North Carolina using tandem mass spectrometry. *MMWR Morb and Mortal Wkly Rep* 2001;**50** (RR-3): 28–9.
9. Hoffman G, Litsheim T, Laessig R. Implementation of tandem mass spectrometry in Wisconsin's newborn screening program. *MMWR Morb Mortal Wkly Rep* 2001;**50** (RR-3): 26–7.
10. Ades AE, Walker J, Jones R, Smith I. Coverage of neonatal screening: failure of coverage or failure of information system. *Arch Dis Child* 2001; **84**:476–9.
11. Office for National Statistics. *Key population and vital statistics (KPVS): local and health authority areas – population and vital statistics by area of usual residence in the United Kingdom, 1999*. Series VS No. 26. London: The Stationery Office; 2001.
12. NHS Centre for Reviews and Dissemination. *Undertaking systematic reviews of research on effectiveness: CRD's guidance for those carrying out or commissioning reviews*. 2nd ed. CRD Report No. 4. York: York Publishing Services; 2001.
13. Greenhalgh T, Donald A. *Evidence based health care workbook: understanding research: for individual and group learning*. London: BMJ Publishing; 2000.
14. Jaeschke R, Guyatt GH, Sackett DL. Evidence-Based Medicine Working Group. Users' guides to medical literature. II: How to use an article about a diagnostic test. A: Are the results of the study valid? *JAMA* 1994;**271**:389–91.
15. Jaeschke R, Guyatt GH, Sackett DL. Evidence-Based Medicine Working Group. Users' guides to medical literature. III: How to use an article about a diagnostic test. B: What are the results and will they help me in caring for my patients? *JAMA* 1994;**271**:703–7.
16. Wiley V, Carpenter K, Wilcken B. Newborn screening with tandem mass spectrometry: 12 months' experience in NSW Australia. *Acta Paediatrica* 1999;**88** (Suppl):48–51.
17. Rashed MS, Rahbeeni Z, Ozand PT. Application of electrospray tandem mass spectrometry to neonatal screening. *Semin Perinatol* 1999;**23**:183–93.
18. Roscher A, Liebl B, Fingerhut R, Olgemoller B. Prospective study of MS–MS newborn screening in Bavaria, Germany. Interim results. *J Inherit Metab Dis* 2000;**23**(Suppl 1):4.
19. Lin WD, Wu JY, Lai CC, Tsai FJ, Tsai CH, Lin SP, Niu DM. A pilot study of neonatal screening by electrospray ionization tandem mass spectrometry in Taiwan. *Acta Paediatr Taiwan* 2001;**42**:224–30.
20. Carpenter K, Wiley V, Sim KG, Heath D, Wilcken B. Evaluation of newborn screening for medium chain acyl-CoA dehydrogenase deficiency in 275 000 babies. *Arch Dis Child Fetal Neonatal Ed* 2001;**85**:F105–9.
21. Pourfarzam M, Morris AA, Appleton M, Craft A, Bartlett K. Neonatal screening for medium-chain acyl-CoA dehydrogenase deficiency. *Lancet* 2001; **358**:1063–4.
22. Chace DH, Hillman SL, Vanhove JLK, Naylor EW. Rapid diagnosis of MCAD deficiency: quantitative analysis of octanoylcarnitine and other acylcarnitines in newborn blood spots by tandem mass spectrometry. *Clin Chem* 1997;**43**:2106–13.

23. Andresen BS, Dobrowolski SF, O'Reilly L, Muenzer J, McCandless SE, Frazier DM, *et al*. Medium-chain acyl-CoA dehydrogenase (MCAD) mutations identified by MS/MS-based prospective screening of newborns differ from those observed in patients with clinical symptoms: identification and characterization of a new, prevalent mutation that results in mild MCAD deficiency. *Am J Hum Genet* 2001;**68**:1408–18.

24. Chace DH, Sherwin JE, Hillman SL, Lorey F, Cunningham GC. Use of phenylalanine-to-tyrosine ratio determined by tandem mass spectrometry to improve newborn screening for phenylketonuria of early discharge specimens collected in the first 24 hours. *Clin Chem* 1998; **44**:2405–9.

25. Rashed MS, Bucknall MP, Little D, Awad A, Jacob M, Alamoudi M, *et al*. Screening blood spots for inborn errors of metabolism by electrospray tandem mass spectrometry with a microplate batch process and a computer algorithm for automated flagging of abnormal profiles. *Clin Chem* 1997;**43**: 1129–41.

26. Schulze A, Lindner M, Olgemoller K, Mayatepek E, Hoffmann GF. Outcome study of extended neonatal screening for inborn errors of metabolism by electrospray ionization tandem mass spectrometry. 5th Meeting of the International Society for Neonatal Screening, Genova, Italy, 26–29 June 2002.

27. Wiley V, Hammond J, Carpenter K, Wilcken B. Establishing performance indicators for the use of tandem mass spectrometry in a newborn screening programme. 5th Meeting of the International Society for Neonatal Screening, Genova, Italy, 26–29 June 2002.

28. Clayton PT, Doig M, Ghafari S, Meaney C, Taylor C, Leonard JV, *et al*. Screening for medium chain acyl-CoA dehydrogenase deficiency using electrospray ionisation tandem mass spectrometry. *Arch Dis Child* 1998;**79**:109–15.

29. Chace DH, Naylor EW. Expansion of newborn screening programs using automated tandem mass spectrometry. *Ment Retard Dev Disab Res Rev* 1999; **5**:150–4.

30. Hutchesson AC, Bundey S, Preece MA, Hall SK, Green A. A comparison of disease and gene frequencies of inborn errors of metabolism among different ethnic groups in the West Midlands, UK. *J Med Genet* 1998;**35**:366–70.

31. Poustie VJ, Rutherford P. Dietary interventions for phenylketonuria. The Cochrane Library (Issue 1). Oxford: Update Software; 2002.

32. Poustie VJ, Rutherford P. Tyrosine supplementation for phenylketonuria. The Cochrane Library (Issue 1). Oxford: Update Software; 2002.

33. Agostoni C, Riva E, Biasucci G, Luotti D, Bruzzese MG, Marangoni F, Giovannini M. The effects of n-3 and n-6 polyunsaturated fatty acids on plasma lipids and fatty acids of treated phenylketonuric children. *Prostaglandins Leukot Essent Fatty Acids* 1995;**53**:401–4.

34. Agostoni C, Scaglioni S, Bonvissuto M, Bruzzese MG, Giovannini M, Riva E. Biochemical effects of supplemented long-chain polyunsaturated fatty acids in hyperphenylalaninemia. *Prostaglandins Leukot Essent Fatty Acids* 2001;**64**:111–5.

35. Griffiths P, Paterson L, Harvie A. Neuropsychological effects of subsequent exposure to phenylalanine in adolescents and young adults with early-treated phenylketonuria. *J Intellect Disabil Res* 1995;**39**:365–72.

36. Robinson M, White FJ, Cleary MA, Wraith E, Lam WK, Walter JH. Increased risk of vitamin B12 deficiency in patients with phenylketonuria on an unrestricted or relaxed diet. *J Pediatr* 2000; **136**:545–7.

37. Waisbren SE, Hanley W, Levy HL, Shifrin H, Allred E, Azen C, *et al*. Outcome at age 4 years in offspring of women with maternal phenylketonuria: the Maternal PKU Collaborative Study. *JAMA* 2000;**283**:756–62.

38. Cerone R, Schiaffino MC, Di Stefano S, Veneselli E. Phenylketonuria: diet for life or not? *Acta Paediatr* 1999;**88**:664–9.

39. Fernandes J, Saudubray JM, van den Berghe G. *Inborn metabolic diseases: diagnosis and treatment*. Berlin: Springer; 1995.

40. Mohan N, McKiernan P, Preece MA, Green A, Buckels J, Mayer AD, Kelly DA. Indications and outcome of liver transplantation in tyrosinaemia type 1. *Eur J Pediatr* 1999;**158** (Suppl 2): S49–54.

41. Schulze A, Frommhold D, Hoffmann GF, Mayatepek E. Spectrophotometric microassay for delta-aminolevulinate dehydratase in dried-blood spots as confirmation for hereditary tyrosinemia type I. *Clin Chem* 2001; **47**:1424–9.

42. Hutchesson AC, Hall SK, Preece MA, Green A. Screening for tyrosinaemia type I. *Arch Dis Child Fetal Neonatal Ed* 1996;**74**:F191–4.

43. Laberge C, Grenier A, Valet JP, Morissette J. Fumarylacetoacetase measurement as a mass screening procedure for hereditary tyrosinaemia type I. *Am J Hum Genet* 1990;**47**:325–8.

44. Goulden KJ, Moss MA, Cole DE, Tithecott GA, Crocker JF. Pitfalls in the initial diagnosis of tyrosinaemia: three case reports and a review of the literature. *Clinical Biochemistry* 1987; **20**:207–12.

45. Scriver CR, Beaudet AL, Sly WS, Valle D. *The metabolic and molecular bases of inherited diseases*. 8th ed. New York: McGraw-Hill; 2001.

46. Franken DG, Blom HJ, Boers GH, Tangerman A, Thomas CM, Trijbels FJ. Thiamine (vitamin B1) supplementation does not reduce fasting blood homocysteine concentration in most homozygotes for homocystinuria. *Biochim Biophys Acta* 1996;**1317**:101–4.

47. Wilcken DE, Wilcken B. The natural history of vascular disease in homocystinuria and the effects of treatment. *J Inherit Metab Dis* 1997;**20**:295–300.

48. Yap S, Naughten E. Homocystinuria due to cystathionine beta-synthase deficiency in Ireland: 25 years' experience of a newborn screened and treated population with reference to clinical outcome and biochemical control. *J Inherit Metab Dis* 1998;**21**:738–47.

49. Yap S, Rushe H, Howard PM, Naughten ER. The intellectual abilities of early treated individuals with pyridoxine-nonresponsive homocystinuria due to cystathionine β-synthase deficiency. *J Inherit Metab Dis* 2001;**24**:437–47.

50. Yap S, Boers GH, Wilcken B, Wilcken DE, Brenton DP, Lee PJ, *et al*. Vascular outcome in patients with homocystinuria due to cystathionine *β*-synthase deficiency treated chronically: a multicenter observational study. *Arterioscler Thromb Vasc Biol* 2001;**21**:2080–5.

51. Mudd SH, Levy HL, Kraus JP. Disorders of transsulfuration. In: Scriver CR, Beaudet AL, Sly WS, Valle D, editors. *The metabolic and molecular bases of inherited diseases*. 8th ed. New York: McGraw-Hill; 2001. pp. 2007–49.

52. Levy HL, Albers S. Genetic screening of newborns. *Annu Rev Genomics and Hum Genet* 2000;**1**:139–77.

53. Morris AAM. Neonatal screening for inborn errors of metabolism. *Indian Pediatr* 2000;**37**:1303–6.

54. Peterschmitt MJ, Simmons JR, Levy HL. Reduction of false negative results in screening of newborns for homocystinuria. *N Engl J Med* 1999;**341**:1572–6.

55. Chace DH, Hillman SL, Millington DS, Kahler SG, Adam BW, Levy HL. Rapid diagnosis of homocystinuria and other hypermethioninemias from newborns' blood spots by tandem mass spectrometry. *Clin Chem* 1996;**42**:349–55.

56. Burlina AB, Ogier H, Korall H, Trefz FK. Long-term treatment with sodium phenylbutyrate in ornithine transcarbamylase-deficient patients. *Mol Genet Metab* 2001;**72**:351–5.

57. Maestri NE, Clissold DB, Brusilow SW. Long-term survival of patients with argininosuccinate synthetase deficiency. *J Pediatr* 1995;**127**:929–35.

58. Nicolaides P, Leonard J, Surtees R. Neurological outcome of methylmalonic acidaemia. *Arch Dis Child* 1998;**78**:508–12.

59. van der Meer SB, Poggi F, Spada M, Bonnefont JP, Ogier H, Hubert P, *et al*. Clinical outcome and long-term management of 17 patients with propionic acidaemia. *Eur J Pediatr* 1996; **155**:205–10.

60. Iafolla AK, Thompson RJ, Roe CR. Medium-chain acyl-coenzyme-A dehydrogenase deficiency: clinical course in 120 affected children. *J Pediatr* 1994;**124**:409–15.

61. Touma EH, Charpentier C. Medium chain acyl-CoA dehydrogenase deficiency. *Arch Dis Child* 1992;**67**:142–5.

62. Wilcken B, Hammond J, Silink M. Morbidity and mortality in medium chain acyl coenzyme A dehydrogenase deficiency. *Arch Dis Child* 1994; **70**:410–12.

63. Wilson CJ, Champion MP, Collins JE, Clayton PT, Leonard JV. Outcome of medium chain acyl-CoA dehydrogenase deficiency after diagnosis. *Arch Dis Child* 1999;**80**:459–62.

64. Heptinstall LE, Till J, Wraith JE, Besley GT. Common MCAD mutation in a healthy parent of two affected siblings. *J Inherit Metab Dis* 1995; **18**:638–9.

65. Pollitt RJ, Leonard JV. Prospective surveillance study of medium chain acyl-CoA dehydrogenase deficiency in the UK. *Arch Dis Child* 1998; **79**:116–19.

66. Seddon HR, Gray G, Pollitt RJ, Iitia A, Green A. Population screening for the common G985 mutation causing medium-chain acyl-CoA dehydrogenase deficiency with Eu-labeled oligonucleotides and the DELFIA system. *Clin Chem* 1997;**43**:436–42.

67. Ziadeh R, Hoffman EP, Finegold DN, Hoop RC, Brackett JC, Strauss AW, Naylor EW. Medium chain acyl-CoA dehydrogenase deficiency in Pennsylvania: neonatal screening shows high incidence and unexpected mutation frequencies. *Pediatr Res* 1995;**37**:675–8.

68. Hoffmann GF, Zschocke J. Glutaric aciduria type I: from clinical, biochemical and molecular diversity to successful therapy. *J Inherit Metab Dis* 1999; **22**:381–91.

69. Monavari AA, Naughten ER. Prevention of cerebral palsy in glutaric aciduria type 1 by dietary management. *Arch Dis Child* 2000;**82**:67–70.

70. Collins J, Leonard JV. Natural history of glutaric aciduria type 1. *Arch Dis Child* 2000;**83**:87.

71. Bjugstad KB, Goodman SI, Freed CR. Age at symptom onset predicts severity of motor impairment and clinical outcome of glutaric acidemia type 1. *J Pediatr* 2000;**137**:681–6.

72. Cairns J, Shackley P. Sometime sensitive, seldom specific: a review of the economic of screening. *Health Econ* 1993;**2**:43–53.

73. O'Brien BJ, Drummond MF, Labelle RJ, Willan A. In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in health care. *Med Care* 1994;**32**:150–63.

74. Briggs AH, Gray AM. Handling uncertainty when performing economic evaluations of healthcare interventions. *Health Technol Assess* 1999;**3**(2).

75. McIntosh E, Donaldson C, Ryan M. Recent advances in the methods of cost-effectiveness analysis in healthcare. *Pharmacoeconomics* 1999; **15**:357–67.

76. Manning WG, Fryback DG, Weinstein MC. Reflecting uncertainty in cost-effectiveness analysis. In Gold MR, Russell LB, Seigal JE, Weinstein MC, editors. *Cost-effectiveness in health and medicine*. New York: Oxford University Press; 1996. pp. 247–75.

77. Drummond MF, O'Brien BJ, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes*. 2nd ed. Oxford: Oxford University Press; 1997.

78. Green A, Pollitt RJ. Population newborn screening for inherited metabolic disease: current UK perspectives. *J Inherit Metab Dis* 1999;**22**:572–9.

79. Briggs AH. Handling uncertainty in economic evaluation. In Drummond MF, McGuire A, editors. *Economic evaluation in health care: merging theory with practice*. Oxford: Oxford University Press; 2001. pp. 172–214.

80. Claxton K, Neumann PJ, Araki SS, Weinstein MC. Bayesian value of information analysis: an application to a policy model of Alzheimer's disease. Discussion Papers in Economics (2000/39). Department of Economics and Related Studies, University of York; 2000.

81. Fenwick E, Claxton K, Sculpher M, Briggs A. Improving the efficiency and relevance of health technology assessment: the role of iterative decision analytic modelling. CHE Discussion Papers, University of York; 2000.

82. Addison GM. MCAD screening: the need for progress. *BMJ* 2001; Electronic letter published 31 January 2001 (http://bmj.com/cgi/eletters/322/7278/112#12359) in response to Tanner S, Sharrard M, Cleary M, Walter J, Wraith E, Lee P, *et al*. Screening for medium chain acyl-CoA dehydrogenase deficiency has still not been evaluated. *BMJ* 2001;**322**:112.

83. Curran AL, Sharples PM, White C, Knapp M. Time costs of caring for children with severe disabilities compared with caring for children without disabilities. *Dev Med Child Neurol* 2001; **43**:529–33.

84. Andersson HC, Marble M, Shapira E. Long-term outcome in treated combined methylmalonic acidemia and homocystinemia. *Genet Med* 1999;**1**:146–50.

85. den Boer ME, Wanders RJ, Morris AA, Ijlst L, Heymans HS, Wijburg FA. Long-chain 3-hydroxyacyl-CoA dehydrogenase deficiency: clinical presentation and follow-up of 50 patients. *Pediatrics* 2002;**109**:99–104.

86. Boles RG, Buck EA, Blitzer MG, Platt MS, Cowan TM., Martin SK, *et al*. Retrospective biochemical screening of fatty acid oxidation disorders in postmortem livers of 418 cases of sudden death in the first year of life. *J Pediatr* 1998;**132**:924–33.

87. Hoffmann GF, Athanassopoulos S, Burlina AB, Duran M, de Klerk JB, Lehnert W, *et al*. Clinical course, early diagnosis, treatment, and prevention of disease in glutaryl-CoA dehydrogenase deficiency. *Neuropediatrics* 1996;**27**:115–23.

88. Adams ME, McCall NT, Gray DT, Orza MJ, Chalmers TC. Economic analysis in randomized control trials. *Med Care* 1992;**30**:231–43.

89. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the *British Medical Journal*. *BMJ* 1996;**313**:275–83.

90. Lord J, Thomason MJ, Littlejohns P, Chalmers RA, Bain MD, Addison GM, *et al*. Secondary analysis of economic data: a review of cost–benefit studies of neonatal screening for phenylketonuria. *J Epidemiol Community Health* 1999;**53**:179–86.

91. Wood JC, Magera MJ, Rinaldo P, Seashore MR, Strauss AW, Friedman A. Diagnosis of very long chain acyl-dehydrogenase deficiency from an infant's newborn screening card. *Pediatrics* 2001;**108**:E19.

92. Tanner S, Sharrard M, Cleary M, Walter J, Wraith E, Lee P, *et al*. Screening for medium chain acyl-CoA dehydrogenase deficiency has still not been evaluated. *BMJ* 2001;**322**:112.

93. Netton A, Rees T, Harrison G. Unit costs of health and social care: 2001. Personal Social Services Research Unit, University of Kent at Canterbury; 2002.

94. Department of Health. *Reference costs 2001*. Leeds: NHS Executive; 2002.

95. Beecham J, O'Neill T, Goodman R. Supporting young adults with hemiplegia: services and costs. *Health Soc Care Community* 2001;**9**:51–9.

96. Hutton JL, Colver AF, Mackie PC. Effects of severity of disability on survival in north east England cerebral palsy cohort. *Arch Dis Child* 2000; **83**:468–74.

97. Hutton JL, Pharoah POD. Effects of cognitive, motor and sensory disabilities on survival in cerebral palsy. *Arch Dis Child* 2002;**86**:84–90.

98. Singer RB, Strauss D, Shavelle R. Comparative mortality in cerebral palsy patients in California 1980–1996. *J Insur Med* 1998;**30**:240–6.

99. Jancar J. Increased life expectancy in people with untreated phenylketonuria. *J Intellect Disab Res* 1998;**42**:97–9.

100. Strauss D, Shavelle R. Life expectancy in adults with cerebral palsy. *Dev Med Child Neurol* 1998; **40**:369–75.

101. NHS Executive. Modernisation of pathology services. Health Service Circular (HSC 1999/170). London: Department of Health; 1999.

102. American Academy of Pediatrics. Newborn screening fact sheets (RE9632). *Pediatrics* 1996; **98**:473–501.

103. Dezateux C. Evaluating newborn screening programmes based on dried blood spots: future challenges. *Br Med Bull* 1998;**54**:877–90.

104. Pollitt RJ. Tandem mass spectrometry screening: proving effectiveness. *Acta Paediatr* 1999; **88** (Suppl 432):40–4.

105. Green A. Neonatal screening: current trends and quality control in the United Kingdom. *Rinsho Byori* 1998;**46**:211–16.

106. Ades AE, Walker J, Jones R, Thompson S, Leonard JV, Smith I. Obstacles to timely neonatal screening in North Thames. *J Med Screen* 1998; **5**:183–6.

107. Leonard JV, Dezateux C. Screening for inherited metabolic diseases in newborn infants using tandem mass spectrometry. *BMJ* 2002;**324**:4–5.

108. Kwon C, Farrell PM. The magnitude and challenge of false-positive newborn screening test results. *Arch Pediatr Adolesc Med* 2000;**154**:714–18.

109. Wang SS, Fernhoff PM, Hannon WH, Khoury MJ. Medium chain acyl-CoA dehydrogenase deficiency: human genome epidemiology review. *Genet Med* 1999;**1**:332–9.

110. Clarke JTR, Gates RD, Hogan SE, Barrett M, MacDonald GW. Neuropsychological studies on adolescents with phenylketonuria returned to phenylalanine-restricted diets. *Am J Ment Retard* 1987;**92**(3):255–62.

111. Griffiths P, Smith C, Harvie A. Transitory hyperphenylalaninaemia in children with continuously treated phenylketonuria. *Am J Ment Retard* 1997;**102**(1):27–36.

112. Griffiths P, Ward N, Harvie A, Cockburn F. Neuropsychological outcome of experimental manipulation of phenylalanine intake in treated phenylketonuria. *J Inherit Metab Dis* 1998;**21**(1):29–38.

113. Holtzman NA, Welcher DW, Mellits ED. Termination of restricted diet in children with phenylketonuria: a randomised controlled study. *N Engl J Med* 1975;**293**(22):1121–4.

114. Mudd SH, Skovby F, Levy HL, Pettigrew KD, Wilcken B, Pyeritz RE, *et al*. The natural history of homocystinuria due to cystathionine β-synthase deficiency. *Am J Hum Genet* 1985;**37**:1–31.

This version of HTA monograph volume 8, number 12 does not include the 180 pages of appendices. This is to save download time from the HTA website.

The printed version of this monograph also excludes the appendices.

View/download the appendices (728 kbytes).

Health Technology Assessment Programme

Prioritisation Strategy Group

Members

Chair,
Professor Tom Walley, Director, NHS HTA Programme & Professor of Clinical Pharmacology, University of Liverpool

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, Radcliffe Hospital, Oxford

Dr Ron Zimmern, Director, Public Health Genetics Unit, Strangeways Research Laboratories, Cambridge

HTA Commissioning Board

Members

Programme Director,
Professor Tom Walley, Director, NHS HTA Programme & Professor of Clinical Pharmacology, University of Liverpool

Chair,
Professor Shah Ebrahim, Professor in Epidemiology of Ageing, Department of Social Medicine, University of Bristol, Canynge Hall, Whiteladies Road, Bristol

Deputy Chair,
Professor Jenny Hewison, Professor of Health Care Psychology, Academic Unit of Psychiatry and Behavioural Sciences, University of Leeds School of Medicine, Leeds

Professor Douglas Altman, Professor of Statistics in Medicine, Centre for Statistics in Medicine, Oxford University, Institute of Health Sciences, Cancer Research UK Medical Statistics Group, Headington, Oxford

Professor John Bond, Professor of Health Services Research, Centre for Health Services Research, University of Newcastle, School of Health Sciences, Newcastle upon Tyne

Professor John Brazier, Director of Health Economics, Sheffield Health Economics Group, School of Health & Related Research, University of Sheffield, ScHARR Regent Court, Sheffield

Dr Andrew Briggs, Public Health Career Scientist, Health Economics Research Centre, University of Oxford, Institute of Health Sciences, Oxford

Dr Christine Clark, Medical Writer & Consultant Pharmacist, Cloudside, Rossendale, Lancs
and
Principal Research Fellow, Clinical Therapeutics in the School of Pharmacy, Bradford University, Bradford

Professor Nicky Cullum, Director of Centre for Evidence Based Nursing, Department of Health Sciences, University of York, Research Section, Seebohm Rowntree Building, Heslington, York

Dr Andrew Farmer, Senior Lecturer in General Practice, Department of Primary Health Care, University of Oxford, Institute of Health Sciences, Headington, Oxford

Professor Fiona J Gilbert, Professor of Radiology, Department of Radiology, University of Aberdeen, Lilian Sutton Building, Foresterhill, Aberdeen

Professor Adrian Grant, Director, Health Services Research Unit, University of Aberdeen, Drew Kay Wing, Polwarth Building, Foresterhill, Aberdeen

Professor Alastair Gray, Director, Health Economics Research Centre, University of Oxford, Institute of Health Sciences, Headington, Oxford

Professor Mark Haggard, Director, MRC ESS Team, CBU Elsworth House, Addenbrooke's Hospital, Cambridge

Professor F D Richard Hobbs, Professor of Primary Care & General Practice, Department of Primary Care & General Practice, University of Birmingham, Primary Care and Clinical Sciences Building, Edgbaston, Birmingham

Professor Peter Jones, Head of Department, University Department of Psychiatry, University of Cambridge, Addenbrooke's Hospital, Cambridge

Professor Sallie Lamb, Research Professor in Physiotherapy/Co-Director, Interdisciplinary Research Centre in Health, Coventry University, Coventry

Dr Donna Lamping, Senior Lecturer, Health Services Research Unit, Public Health and Policy, London School of Hygiene and Tropical Medicine, London

Professor David Neal, Professor of Surgical Oncology, Oncology Centre, Addenbrooke's Hospital, Cambridge

Professor Tim Peters, Professor of Primary Care Health Services Research, Division of Primary Health Care, University of Bristol, Cotham House, Cotham Hill, Bristol

Professor Ian Roberts, Professor of Epidemiology & Public Health, Intervention Research Unit, London School of Hygiene and Tropical Medicine, London

Professor Peter Sandercock, Professor of Medical Neurology, Department of Clinical Neurosciences, University of Edinburgh, Western General Hospital NHS Trust, Bramwell Dott Building, Edinburgh

Professor Martin Severs, Professor in Elderly Health Care, Portsmouth Institute of Medicine, Health & Social Care, St George's Building, Portsmouth

Dr Jonathan Shapiro, Senior Fellow, Health Services Management Centre, Park House, Birmingham

Current and past membership details of all HTA 'committees' are available from the HTA website (www.ncchta.org)

Diagnostic Technologies & Screening Panel

Members

Chair,
Dr Ron Zimmern, Director of the Public Health Genetics Unit, Strangeways Research Laboratories, Cambridge

Dr Paul Cockcroft, Consultant Medical Microbiologist/ Laboratory Director, Public Health Laboratory, St Mary's Hospital, Portsmouth

Professor Adrian K Dixon, Professor of Radiology, Addenbrooke's Hospital, Cambridge

Dr David Elliman, Consultant in Community Child Health, London

Dr Andrew Farmer, Senior Lecturer in General Practice, Institute of Health Sciences, University of Oxford

Dr Karen N Foster, Clinical Lecturer, Dept of General Practice & Primary Care, University of Aberdeen

Professor Jane Franklyn, Professor of Medicine, University of Birmingham

Professor Antony J Franks, Deputy Medical Director, The Leeds Teaching Hospitals NHS Trust

Mr Tam Fry, Honorary Chairman, Child Growth Foundation, London

Dr Susanne M Ludgate, Medical Director, Medical Devices Agency, London

Dr William Rosenberg, Senior Lecturer and Consultant in Medicine, University of Southampton

Dr Susan Schonfield, CPHM Specialised Services Commissioning, Croydon Primary Care Trust

Dr Margaret Somerville, Director of Public Health, Teignbridge Primary Care Trust, Devon

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

Dr Dennis Wright, Consultant Biochemist & Clinical Director, Pathology & The Kennedy Galton Centre, Northwick Park & St Mark's Hospitals, Harrow

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 and Cochrane Editor, Bushey Health Centre, Bushey, Herts.

Mr Charles Dobson, Special Projects Adviser, Department of Health

Dr Robin Ferner, Consultant Physician and Director, West Midlands Centre for Adverse Drug Reactions, City Hospital NHS Trust, Birmingham

Dr Karen A Fitzgerald, Pharmaceutical Adviser, Bro Taf Health Authority, Cardiff

Professor Alastair Gray, Professor of Health Economics, 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

Professor Robert Peveler, Professor of Liaison Psychiatry, Royal South Hants Hospital, Southampton

Dr Frances Rotblat, CPMP Delegate, Medicines Control Agency, London

Mrs Katrina Simister, New Products Manager, National Prescribing Centre, Liverpool

Dr Ken Stein, Senior Lecturer in Public Health, University of Exeter

Professor Terence Stephenson, Professor of Child Health, University of Nottingham

Dr Richard Tiner, Medical Director, Association of the British Pharmaceutical Industry, London

Professor Dame 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 and General Surgeon, Royal Devon & Exeter Hospital

Dr Mahmood Adil, Head of Clinical Support & Health Protection, Directorate of Health and Social Care (North), Department of Health, Manchester

Professor John Bond, Head of Centre for Health Services Research, University of Newcastle upon Tyne

Mr Michael Clancy, Consultant in A & E Medicine, Southampton General Hospital

Dr Carl E Counsell, Senior Lecturer in Neurology, University of Aberdeen

Dr Keith Dodd, Consultant Paediatrician, Derbyshire Children's Hospital, Derby

Professor Gene Feder, Professor of Primary Care R&D, Barts & the London, Queen Mary's School of Medicine and Dentistry, University of London

Ms Bec Hanley, Freelance Consumer Advocate, Hurstpierpoint, West Sussex

Professor Alan Horwich, Director of Clinical R&D, The Institute of Cancer Research, London

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

Dr Mike McGovern, Senior Medical Officer, Heart Team, Department of Health, London

Dr John C Pounsford, Consultant Physician, North Bristol NHS Trust

Professor Mark Sculpher, Professor of Health Economics, Institute for Research in the Social Services, University of York

Dr L David Smith, Consultant Cardiologist, Royal Devon & Exeter Hospital

Professor Norman Waugh, Professor of Public Health, University of Aberdeen

Expert Advisory Network

Members

Mr Gordon Aylward,
Chief Executive,
Association of British Health-Care Industries, London

Ms Judith Brodie,
Head of Cancer Support Service, Cancer BACUP, London

Mr Shaun Brogan,
Chief Executive, Ridgeway Primary Care Group, Aylesbury, Bucks

Ms Tracy Bury,
Project Manager, World Confederation for Physical Therapy, London

Mr John A Cairns,
Professor of Health Economics, Health Economics Research Unit, University of Aberdeen

Professor Howard Stephen Cuckle,
Professor of Reproductive Epidemiology, Department of Paediatrics, Obstetrics & Gynaecology, University of Leeds

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, London

Professor Martin Eccles,
Professor of Clinical Effectiveness, Centre for Health Services Research, University of Newcastle upon Tyne

Professor Pam Enderby,
Professor of Community Rehabilitation, Institute of General Practice and Primary Care, University of Sheffield

Mr Leonard R Fenwick,
Chief Executive, Newcastle upon Tyne Hospitals NHS Trust

Professor David Field,
Professor of Neonatal Medicine, Child Health, The Leicester Royal Infirmary NHS Trust

Mrs Gillian Fletcher,
Antenatal Teacher & Tutor and President, National Childbirth Trust, Henfield, West Sussex

Ms Grace Gibbs,
Deputy Chief Executive, Director for Nursing, Midwifery & Clinical Support Servs., West Middlesex University Hospital, Isleworth, Middlesex

Dr Neville Goodman,
Consultant Anaesthetist, Southmead Hospital, Bristol

Professor Robert E Hawkins,
CRC Professor and Director of Medical Oncology, Christie CRC Research Centre, Christie Hospital NHS Trust, Manchester

Professor F D Richard Hobbs,
Professor of Primary Care & General Practice, Department of Primary Care & General Practice, University of Birmingham

Professor Allen Hutchinson,
Director of Public Health & Deputy Dean of ScHARR, Department of Public Health, University of Sheffield

Professor Rajan Madhok,
Medical Director & Director of Public Health, Directorate of Clinical Strategy & Public Health, North & East Yorkshire & Northern Lincolnshire Health Authority, York

Professor David Mant,
Professor of General Practice, Department of Primary Care, University of Oxford

Professor Alexander Markham,
Director, Molecular Medicine Unit, St James's University Hospital, Leeds

Dr Chris McCall,
General Practitioner, The Hadleigh Practice, Castle Mullen, Dorset

Professor Alistair McGuire,
Professor of Health Economics, London School of Economics

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

Professor Jon Nicholl,
Director of Medical Care Research Unit, School of Health and Related Research, University of Sheffield

Mrs Julietta Patnick,
National Co-ordinator, NHS Cancer Screening Programmes, Sheffield

Professor Chris Price,
Visiting Chair – Oxford, Clinical Research, Bayer Diagnostics Europe, Cirencester

Ms Marianne Rigge,
Director, College of Health, London

Professor Sarah Stewart-Brown,
Director HSRU/Honorary Consultant in PH Medicine, Department of Public Health, University of Oxford

Professor Ala Szczepura,
Professor of Health Service Research, Centre for Health Services Studies, University of Warwick

Dr Ross Taylor,
Senior Lecturer, Department of General Practice and 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.

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