A systematic review of evidence for the appropriateness of neonatal screening programmes for inborn errors of metabolism

Margaret J. Thomason, Joanne Lord, Murray D. Bain, Ronald A. Chalmers, Peter Littlejohns, G. Michael Addison, A. Hervey Wilcox and Carol A. Seymour

Abstract

Background Developments in screening technology and increased understanding of the natural history and treatment of inborn errors of metabolism (IEMs) have produced pressure to extend neonatal screening programmes. This review aims to assess the evidence for the appropriateness of such programmes.

Methods A formal systematic literature review was conducted. Exclusion and inclusion criteria were used to select papers for critical appraisal by pairs of reviewers. Standard criteria were used to assess the appropriateness of neonatal screening for various IEMs. Site visits were conducted to assess new technologies for newborn screening.

Results A total of 1866 papers were identified and 407 systematically selected for full critical appraisal. Published evidence confirmed that universal newborn screening for phenylketonuria (PKU) meets all of the screening criteria and justifies the expense and infrastructure necessary for the collection and testing of neonatal blood spots. There was insufficient evidence in the literature to assess the cost-effectiveness of screening for any other IEMs. There was reasonable evidence to support inclusion in extended neonatal screening of four other IEMs: biotinidase deficiency, congenital adrenal hyperplasia (CAH), medium-chain acyl CoA dehydrogenase (MCAD) deficiency and glutaric aciduria type 1 (GA1).

Conclusions Large-scale trials of screening for biotinidase, CAH, MCAD and GA1 should be conducted, with careful evaluation to establish their clinical effectiveness and cost-effectiveness in practice. Screening for the latter two disorders would be dependent upon the use of tandem mass spectrometry (tandem MS). The application of tandem MS to newborn screening requires further evaluation. The extension of neonatal screening programmes to other IEMs is not currently justified.

Keywords: neonatal screening, inborn errors of metabolism, mass spectrometry, literature review

Introduction

The first population-based neonatal screening programmes for detecting inborn errors of metabolism (IEMs) were introduced around 30 years ago for phenylketonuria (PKU). Since that time there has been a substantial increase in the number of identified IEMs and although each individual disease is rare the total incidence is significant. For example, in the United Kingdom, with around 793,000 infants born each year, there are 60–70 affected by PKU, 60 infants with congenital adrenal hyperplasia (CAH), 50–60 with medium-chain acyl CoA dehydrogenase (MCAD) deficiency and 40–50 with disorders of organic acid metabolism. It has been estimated that these disorders collectively have an impact on health care comparable with that of juvenile-onset diabetes.

More than 99 per cent of newborns in the United Kingdom, and infants in most other developed countries, are currently tested for PKU and also for congenital hypothyroidism. There is considerable pressure to expand universal newborn screening programmes to cover a broader range of IEMs. This has been further supported by the development of new technologies which are applicable to multi-disease screening programmes, for example tandem mass spectrometry (tandem MS) and...
automated immunoassay. Before the introduction of any such screening programme there should be good evidence that the additional screening and consequent treatment costs, and the potential anxiety to parents, are outweighed by the benefits of health improvements and savings in health care. The fact that a wide range of metabolic disorders can be detected by using new technologies does not necessarily mean that screening for these disorders should be undertaken. The longer-term implications for individuals, the health service and society should also be considered. For some IEM disorders it is not even clear that early detection offers any benefit to the child.3

A health technology assessment project was undertaken in the United Kingdom to assess current newborn screening programmes, to evaluate the available evidence for their expansion and to make recommendations for future developments. Two fundamental questions were addressed: (1) Which IEMs should be screened for? (2) What technologies should be used for screening?

A consortium was established to review the technical, clinical and economic implications of screening. The project team consisted of a combination of the disciplines of clinical biochemistry, metabolic medicine (both adult and paediatric), public health and health economics. Two members of this team were also experts on the automation of technologies and on mass spectrometry. Although the aim was to provide advice to the National Health Service the literature review was international and the results are pertinent to those countries with comparable incidence rates of IEMs to those in the United Kingdom. To avoid compromising the high rate of coverage of current newborn screening programmes the assumption was made that the infrastructures for sample collection would not be substantially altered.

Methods

Three methods of data collection were used: a formal systematic review of the literature was conducted, a questionnaire was sent to all newborn screening laboratories in the United Kingdom to enquire about current practice and grey literature, and visits were made in the United Kingdom, United States and Finland to assess potential new technologies for newborn screening.

For the purpose of the systematic review a classical definition of an IEM was taken, i.e. a monogenic disease resulting in deficient activity in a single enzyme in a pathway of intermediary metabolism. Disorders thus considered were the catabolic and synthetic pathways of carbohydrates, amino acids, organic acids, fatty acids, purines, pyrimidines, porphyrins, steroids, lipids and bile acids, and secondly of the processes involved in the uptake, synthesis and utilization of the essential cofactors for enzymes in these pathways, including trace metals and vitamins; also included were disorders of lysosomal enzymes. The categories of IEM included in the review were:

(1) phenylketonuria;
(2) amino acidopathies;
(3) disorders of carbohydrate metabolism;
(4) disorders of organic acid metabolism;
(5) fatty acid oxidation defects;
(6) disorders of adrenal steroidogenesis;
(7) lipoprotein disorders;
(8) peroxisomal disorders;
(9) disorders of the urea cycle;
(10) respiratory chain/tcicarboxylic acid cycle disorders;
(11) trace metal disorders;
(12) purine–pyrimidine disorders;
(13) lysosomal disorders.

Other inherited disorders, for example of membrane transport, of connective tissue, of blood and of blood-forming tissues, of the defence and immune systems, and of muscle and skin were excluded. Congenital hypothyroidism, which can be caused by metabolic defects of thyroxine synthesis, was not considered.

The search methods and selection criteria are shown in Fig. 1. A total of 407 papers were selected for critical appraisal. For each of the 13 categories of IEM at least two subject experts read and critically appraised all of the selected papers using an agreed checklist, which included study population, screening method, incidence, clinical follow-up procedure, outcomes and costs. Conclusions regarding each disorder were drawn in relation to the fulfilment of seven criteria based on the Wilson and Jungner screening criteria.6

(1) the disorder should be clinically and biochemically well defined;
(2) there should be a known incidence of the disease in populations relevant to the United Kingdom;
(3) the disorder should be associated with significant morbidity or mortality;
(4) effective treatment should be available;
(5) there should be a period before the onset of the disease during which intervention improves outcome;
(6) there should be an ethical safe, simple and robust screening test for the disease;
(7) the screening should be cost effective.

Evidence on the last criterion (cost-effectiveness) was available only in very few studies and was reviewed separately, as were papers relating to multi-disease screening technologies. No published controlled intervention studies were identified; the literature consisted of uncontrolled trials of screening and observational studies.7 There were no data suitable for meta-analysis. As all the literature was grade III or less,6 inclusion criteria were content based only.

Results

Disease-based appraisals

The fulfilment of the screening criteria for those IEMs which are either currently screened for, or for which newborn
Figure 1 Methods of identification and selection of papers.
screening has been suggested in the literature, are listed in Tables 1–4. Ideally, for a disease to be identified as suitable for inclusion in a neonatal screening programme, all the criteria should be fulfilled. (Because of the lack of cost-effectiveness data for any disorders other than PKU this criterion is omitted from the tables.)

Table 1 shows that PKU fulfils all the criteria and may therefore be taken as the standard against which to compare other disorders. Non-PKU aminoacidopathies, because of their uncertain incidence and a lack of evidence of effective treatment, do not fulfil all the criteria. Among the disorders of fatty acid oxidation (Table 2) medium-chain acyl CoA dehydrogenase (MCAD) deficiency shows greatest fulfilment of the criteria, because of its high incidence and the potential large health benefits from treatment, although the natural history of this disorder is not yet fully known and it has been suggested that many individuals may remain asymptomatic. Other disorders of fatty acid oxidation fail significantly to fulfil the screening criteria. Screening for MCAD deficiency is dependent upon the use of tandem MS. None of the ‘common’ organic acid disorders (Table 3) show even moderately good fulfilment of the screening criteria, many presenting acutely in the newborn period and most showing relatively poor long-term outcome despite improvements in therapy. There is additionally no evidence to suggest that early detection and treatment improve outcome. Glutaric aciduria type 1 (GA1) may be an exception, as this disorder has a significant asymptomatic period and is associated with severe neurological sequelae that may be totally preventable by simple treatment. Similar factors relate to biotinidase deficiency (Table 4) despite the relatively low incidence of the complete deficiency. In both cases (GA1 and biotinidase) therapy is simple, effective and cheap. Screening for GA1 is dependent upon the use of tandem MS whereas screening for biotinidase deficiency utilizes a simple assay of enzyme activity.

Congenital adrenal hyperplasia (CAH) caused by deficiency of the enzyme 21-hydroxylase (Table 4) would completely fulfill the screening criteria except that for most females there is essentially no asymptomatic period; these infants should be identified clinically at birth or at the post-natal examination. However, there is still benefit from early detection and diagnosis from newborn screening by avoiding diagnostic uncertainty and preventing gender mis-assignment. Newborn screening programmes for CAH have recently been introduced in both France and Sweden.

Previous justification for neonatal screening programmes for galactosaemia (Table 4) has been based on prevention of morbidity; however, published evidence suggests that, despite early treatment, long-term outcome is poor, with a continuing degree of neurological handicap. Screening for heterozygous familial hypercholesterolaemia (Table 4) has been proposed to allow treatment for this condition to be started before the development of symptomatic disease. However, measurement of total cholesterol or apolipoproteins in the neonatal period is poorly predictive of later values and no proven benefits have yet been shown to arise from treatment of children very early in life. Similarly, screening for Wilson disease (Table 4) by the measurement of ceruloplasmin, the copper transporting protein of plasma, has been proposed but its concentration in neonates with Wilson disease is not yet known. Thus newborn screening programmes are not currently recommended for any of these three IEMs.

Review of screening technologies

Several different techniques including bacterial inhibition assays (Guthrie), chromatography and fluorimetry are currently used to screen for PKU. Dried blood spots are generally used but some centres use liquid blood samples. A laboratory screening for more than one disorder will normally use more than one technique and the methods used vary from manual to partly automated. Full automation of neonatal screening, whatever technique is used, will require the development of an automated punch capable of assessing the quality and position of blood spots on the filter paper.

The technology already exists which could lead to the development of a fully automated system capable of performing several tests. By the use of different labels, time-resolved fluorescence, currently used for congenital hypothyroidism screening, can measure up to four analytes simultaneously. This number of tests can be further widened by the use of a detector with photometric and luminometric capabilities in addition to fluorimetry; such an instrument can combine immunoassay with enzymatic and chemical methods. Such machines will be expensive (more than £100,000) and would only be likely to be cost effective in large-scale screening laboratories.

Molecular (DNA) techniques appear generally unsuitable for universal neonatal screening for IEMs, as these disorders are almost all caused by a very large number of different individual mutations, sometimes specific to families, and genotype-phenotype correlations are not clear. The ideal role of molecular testing in newborn screening may be for confirmation in cases identified from more conventional methods, especially where the latter are less specific (e.g. for cystic fibrosis).

Tandem MS has the potential for simultaneous multi-disease screening, including PKU, using a single analytical technique and would be complementary to immunoassay-based methods required for congenital hypothyroidism, cystic fibrosis and CAH screening. Evidence on the current status and future of potential neonatal screening by tandem MS was obtained both from the literature review and from visits to four laboratories, two in the United Kingdom and two in the United States, currently utilizing or developing the technique. The procedure used was similar in all the laboratories and all use a punched sample from the dried blood spot card. The technology has been demonstrated to be robust (i.e. sensitive and specific) and suitable for the reliable detection of PKU and certain other IEMs. However, concrete data from prospective newborn
Table 1 Fulfilment of screening criteria for selected amino acid disorders

<table>
<thead>
<tr>
<th>Amino acid disorder</th>
<th>Clinically and biochemically well-defined disorder</th>
<th>Known incidence in populations relevant to UK</th>
<th>Associated with significant morbidity or mortality</th>
<th>Effective treatment available</th>
<th>Period before onset during which intervention improves outcome</th>
<th>Ethical, safe, simple and robust screening test</th>
</tr>
</thead>
<tbody>
<tr>
<td>PKU</td>
<td>✔ Clinical and biochemical spectrum of disease well known(^{31})</td>
<td>✔ 1:12 000 (average UK figure(^{1}))</td>
<td>✔ Severe neurological damage impairing cognitive development; early death</td>
<td>✔ Dietary therapy(^{32})</td>
<td>✔ Early diagnosis and treatment reduce incidence of neurological handicap from 80–90% to 6–8%(^{32})</td>
<td>✔ Dried blood spots tested with bacterial inhibition test, chromatography or fluorometry. Tandem MS(^{16}) can also be used</td>
</tr>
<tr>
<td>Maple syrup urine disease (MSUD)</td>
<td>✔ Clinical and biochemical phenotypes well described(^{33})</td>
<td>? Average 1:185 000 world-wide;(^{33}) 0:467 448 screened in Scotland(^{34})</td>
<td>✔ Developmental delay, seizures and encephalopathy. Presentation in neonatal period usually fatal without treatment and later in life death may occur during episodes of metabolic development</td>
<td>? Dietary therapy, but does not completely prevent metabolic crises(^{36}) or prevent developmental delay</td>
<td>? Neonatal onset usually at 4–7 days (possible before result of screening available)</td>
<td>✔ Dried blood spots tested with bacterial inhibition test.(^{34}) Tandem MS(^{36}) can also be used</td>
</tr>
<tr>
<td>Homocystinuria</td>
<td>✔ Natural history now well delineated(^{37})</td>
<td>? Average 1:291 000 world-wide(^{37}) Estimated 1:234 000 in UK, but significant regional variation</td>
<td>✔ Neurological dysfunction, thrombo-embolic events, impaired vision and ectopia lentis</td>
<td>? Pyridoxine and dietary therapy with methionine restriction and L-cysteine supplementation</td>
<td>✔ Early diagnosis and treatment improves IQ by some 35 points(^{37})</td>
<td>✔ Dried blood spots for associated hypermethioninemia; bacterial inhibition assay(^{40}) or chromatography unreliable but tandem MS much more robust(^{41})</td>
</tr>
<tr>
<td>Tyrosinaemia type 1</td>
<td>✔ Spectrum of clinical presentations well characterized(^{38}) together with associated biochemical phenotype</td>
<td>? 1:105 000(^{29})</td>
<td>✔ Liver failure, renal tubular dysfunction, neuropsychiatric crises and hepatocellular carcinoma during and after second decade of life</td>
<td>? Dietary therapy effective in short term, but alternative treatments eventually needed, e.g liver transplant. Long-term efficacy of NTBC treatment still under investigation</td>
<td>? Acute onset within weeks of birth or delayed into infancy and childhood</td>
<td>✔ Dried blood spots by fluorometry.(^{42}) Also tandem MS(^{16})</td>
</tr>
</tbody>
</table>

✔ Criteria fulfilled.
? More data required.
X Criteria not fulfilled.
Table 2 Fulfilment of screening criteria for selected disorders of fatty acid β-oxidation

<table>
<thead>
<tr>
<th>Fatty acid disorder</th>
<th>Clinically and biochemically well-defined disorder</th>
<th>Known incidence in populations relevant to UK</th>
<th>Associated with significant morbidity or mortality</th>
<th>Effective treatment available</th>
<th>Period before onset during which intervention improves outcome</th>
<th>Ethical, safe, simple and robust screening test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medium-chain acyl CoA dehydrogenase (MCAD) deficiency</td>
<td>✓ Genetically and clinically fairly well characterized. Incidence and long-term outcome unknown for asymptomatic individuals</td>
<td>? 1:6500 to 1:20 000 (estimated)[^43]</td>
<td>✓ Life-threatening hypoglycaemic encephalopathy from neonatal period. Acute rhabdomyolysis in adults now also recognized</td>
<td>✓ Regular high-carbohydrate intake, emergency dietary regimen for illness, L-carnitine therapy</td>
<td>✓ Although presentation in newborn period now described, large majority are asymptomatic during the period of newborn screening</td>
<td>✓ Dried blood spots; determination of octanoylcarnitine using tandem MS[^43]</td>
</tr>
<tr>
<td>Mitochondrial long-chain fatty acid oxidation disorders</td>
<td>X Clinical and biochemical features variable and diverse. Outcome poor to unknown</td>
<td>X Unknown but appears rare</td>
<td>? Hypoglycaemia, encephalopathy, early death in some cases. Cardiomyopathy in older patients; others less severe disease</td>
<td>? Treatment of milder cases may be effective but outcome in severe cases is poor</td>
<td>X Not known for all conditions and presentations</td>
<td>? Dried blood spots; determination of long-chain acylcarnitines (C(<em>{12})-C(</em>{18})) using tandem MS but these are currently difficult to detect and not all cases may be identified</td>
</tr>
<tr>
<td>Multiple acyl CoA dehydrogenase deficiency (MACDD)</td>
<td>X Presentation highly variable and natural history unknown</td>
<td>X Unknown but appears rare</td>
<td>? Presentation ranges from death in infancy with acidosis, hypoglycaemia, hyperammonaemia, convulsions and neurological symptoms to much milder disease in later childhood/adult life</td>
<td>? Milder cases may respond to an increased carbohydrate and reduced fat and protein intake, riboflavin and L-carnitine. Severe infantile disease often refractory to therapy</td>
<td>X Severe neonatal disease often presents before screening result would be available. Associated dysmorphic features and polycystic kidneys in some cases imply intrauterine damage</td>
<td>✓ Dried blood spots; determination of relevant acylcarnitines using tandem MS possible but not yet proven</td>
</tr>
</tbody>
</table>

\[^43\]References throughout.
✓ Criteria fulfilled.
? More data required.
X Criteria not fulfilled.
<table>
<thead>
<tr>
<th>Organic acid disorder</th>
<th>Clinically and biochemically well-defined disorder</th>
<th>Known incidence in populations relevant to UK</th>
<th>Associated with significant morbidity or mortality</th>
<th>Effective treatment available</th>
<th>Period before onset during which intervention improves outcome</th>
<th>Ethical, safe, simple and robust screening test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methylmalonic aciduria</td>
<td>? Generally well characterized and defined although long-term outcome still unclear</td>
<td>? Not known with certainty but probably &lt;1:80 000</td>
<td>✓ Life-threatening episodes of acute illness with longer-term problems of neurological handicap and renal damage</td>
<td>? Improving with diet, antibiotics, L-carnitine, somatotrophin, organ transplantation, and B12 in vitamin responsive cases</td>
<td>? Many cases present in neonatal period, others later in infancy and childhood. No evidence early intervention improves outcome</td>
<td>✓ Dried blood spots; determination of propionylcarnitines using tandem MS but levels may be very low in some cases</td>
</tr>
<tr>
<td>Propionic acidaemia</td>
<td>? Reasonably well characterized and defined. Long-term outcome in 'milder' cases unknown</td>
<td>? Probably &lt;1:100 000</td>
<td>✓ Life-threatening episodes of acute illness with longer-term problems of neurological handicap</td>
<td>? Improving with diet, antibiotics, L-carnitine, somatotrophin, liver transplantation, B12 in vitamin responsive cases</td>
<td>? Many cases present in neonatal period. Evidence of improved outcome following early intervention not available</td>
<td>✓ Dried blood spots; determination of propionylcarnitines using tandem MS</td>
</tr>
<tr>
<td>Isovaleric acidaemia</td>
<td>✓ Well characterized</td>
<td>? Very low in UK, probably &lt;1:200 000</td>
<td>✓ Presents with acidosis, vomiting, tremors, coma and death in the neonatal period and beyond</td>
<td>✓ Diet plus glycine and L-carnitine therapy</td>
<td>? Many cases present acutely in neonatal period. No evidence early intervention improves outcome but will prevent acute episodes</td>
<td>✓ Dried blood spots; determination of isovaleryl carnitines using tandem MS</td>
</tr>
<tr>
<td>Glutaric aciduria type 1</td>
<td>? Becoming better characterized but complete natural history unknown</td>
<td>? Unclear but may be as high as 1:50 000</td>
<td>✓ Severe crippling choreoathetoid cerebral palsy of postnatal onset</td>
<td>? Current opinion suggests L-carnitine therapy presymptomatically may prevent onset of neurological damage</td>
<td>✓ All patients seem to have an asymptomatic period of a month or more</td>
<td>✓ Dried blood spots; determination of glutaryl carnitines by tandem MS</td>
</tr>
<tr>
<td>3-hydroxy-3-methylglutaric aciduria</td>
<td>? Long-term outcome with therapy unknown</td>
<td>? Not known, probably &lt;1:80 000</td>
<td>✓ Episodic hypokekapetic hypoglycaemia causing coma and death and possible long-term neurological sequelae into at least teenage years</td>
<td>✓ Moderate dietary modification, emergency regimen for illness, and L-carnitine therapy. Need for lifelong therapy unknown</td>
<td>? Generally there is asymptomatic period though neonatal presentation is recognized</td>
<td>✓ Dried blood spots; determination of 3-methyl glutaryl carnitines using tandem MS</td>
</tr>
</tbody>
</table>

References throughout.

✓ Criteria fulfilled.

? More data required.

X Criteria not fulfilled.
<table>
<thead>
<tr>
<th>Metabolic disorder</th>
<th>Clinically and biochemically well-defined disorder</th>
<th>Known incidence in populations relevant to UK</th>
<th>Associated with significant morbidity or mortality</th>
<th>Effective treatment available</th>
<th>Period before onset during which intervention improves outcome</th>
<th>Ethical, safe, simple and robust screening test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biotinidase deficiency</td>
<td>✔ Causes multiple carboxylase deficiency. Long-term outcome of therapy to be fully evaluated</td>
<td>? Estimated in UK about 1:110 000\textsuperscript{66}</td>
<td>✔ Progressive neurological disease which may be fatal in many cases</td>
<td>✔ Oral biotin. Requirements for life-long therapy not known</td>
<td>✔ Presentation generally in early infancy and childhood</td>
<td>✔ Dried blood spots; colorimetric assay of enzyme activity\textsuperscript{51} Need to distinguish subjects with partial deficiency</td>
</tr>
<tr>
<td>Congenital adrenal hyperplasia (CAH)</td>
<td>✔ Well defined. Refers to 21-hydroxylase deficiency</td>
<td>✔ Estimated 1:6000 to 1:21 000\textsuperscript{59,50}</td>
<td>✔ Hypogonadism, dehydration, virilization</td>
<td>✔ Steroid replacement</td>
<td>? Clinically present at 3 days to 8 weeks, most commonly at 3 weeks</td>
<td>✔ Determination of 17α-hydroxy progesterone from dried blood spots\textsuperscript{51}</td>
</tr>
<tr>
<td>Familial hypercholesterolaemia (FH)</td>
<td>✔ Well-characterized cause of hypercholesterolaemia</td>
<td>✔ Homozygotes 1:1 000 000\textsuperscript{52} Heterozygotes 1:200 to 1:1000</td>
<td>✔ Heterozygotes: vascular damage (atherosclerosis) estimated to cause 5% of premature coronary heart disease (CHD)\textsuperscript{53} Approx. 50% men with FH develop CHD by age 50, 50% women by 60 Heterozygotes: fatal coronary heart disease only angiographically detectable in young adult life (heterozygotes)</td>
<td>? Benefit from intervention (heterozygotes) during childhood yet to be demonstrated</td>
<td>✔ Coronary heart disease only angiographically detectable in young adult life (heterozygotes)</td>
<td>X Total or LDL cholesterol or apolipoprotein B can be measured from blood spots\textsuperscript{55} but are poorly predictive of later values</td>
</tr>
<tr>
<td>Galactosaemia</td>
<td>✔ Very well-defined defect</td>
<td>✔ 1:44 000\textsuperscript{13}</td>
<td>✔ Liver and renal damage, cataract formation, immunodeficiency, neurological damage</td>
<td>✔ Dietary therapy reverses hepatotoxicity and prevents cataract development but not progressive neurodegenerative disorder\textsuperscript{56} or ovarian failure</td>
<td>? Present typically in second week of life</td>
<td>? Colorimetric enzyme assay on dried blood spots\textsuperscript{60} False positive rate 0.03%, false negative rate 0–7%. Beutler fluorescent assay\textsuperscript{61}</td>
</tr>
<tr>
<td>Wilson disease</td>
<td>✔ Well-characterized defect in copper transport</td>
<td>✔ 1:50 000 to 1:100 000\textsuperscript{55}</td>
<td>✔ Liver failure neurological damage</td>
<td>✔ Penicillamine. Pre-symptomatic treatment with zinc may be possible\textsuperscript{57}</td>
<td>✔ Most cases present in late childhood, adolescence and early adulthood\textsuperscript{58,59}</td>
<td>✔ Enzymatic assay of caeruloplasmin\textsuperscript{62} or ELISA method\textsuperscript{63} but not yet validated for use in newborns</td>
</tr>
</tbody>
</table>

✔ Criteria fulfilled.
? More data required.
X Criteria not fulfilled.
Table 5 Collective fulfilment of criteria and recommendations for newborn screening

<table>
<thead>
<tr>
<th>Inborn error of metabolism</th>
<th>Fulfilment of screening criteria and recommendation for newborn screening</th>
</tr>
</thead>
<tbody>
<tr>
<td>PKU</td>
<td>(1) Fulfils all screening criteria (including cost-effectiveness)</td>
</tr>
<tr>
<td>Maple syrup urine disease (MSUD)</td>
<td>(3) Low incidence, neonatal onset often pre-dates availability of screening result, long-term outcome of treatment variable.</td>
</tr>
<tr>
<td>Homocystinuria</td>
<td>(3) Low incidence and effectiveness of treatment over a lifetime unproven</td>
</tr>
<tr>
<td>Tyrosinaemia type 1</td>
<td>(3) Low incidence, no evidence of improved outcome from early, pre-symptomatic treatment</td>
</tr>
<tr>
<td>Medium-chain acyl CoA dehydrogenase (MCAD) deficiency</td>
<td>(2) High frequency, morbidity and mortality largely preventable by simple, cheap and effective treatment, simple screening test requires tandem MS. Screening research programme required to assess exact incidence, long-term outcomes and cost-effectiveness</td>
</tr>
<tr>
<td>Mitochondrial long-chain fatty acid oxidation disorders</td>
<td>(4) Unknown natural history, low incidence, ineffective treatments in many cases and possible ambiguous screening results with possible false negatives using tandem MS</td>
</tr>
<tr>
<td>Multiple acyl CoA dehydrogenase deficiency (MACDD)</td>
<td>(4) Widely variable clinical presentation, some cases with intrauterine damage, dysmorphology, acute neonatal presentation, ineffective treatment, low unknown incidence and unproven screening test (although possible with tandem MS)</td>
</tr>
<tr>
<td>Methylmalonic aciduria } Propionic acidemia</td>
<td>(3) Highly variable outcome, many cases present in newborns before screening result available, treatment efficacy variable. At best, potential candidates for newborn screening. Incidence has to be considered collectively with other similar disorders detectable using tandem MS. Tandem MS does not distinguish these two organic acid disorders</td>
</tr>
<tr>
<td>Isovaleric acidemia</td>
<td>(3) Very low incidence offset by collective incidence of other disorders; early treatment may prevent acute episodes and prevent neurological damage. Potential candidate for newborn screening, requires tandem MS</td>
</tr>
<tr>
<td>Glutaric aciduria type 1*</td>
<td>(3) Potentially preventable severe neurological and disabling disorder, thus a candidate for newborn screening despite probable low incidence. Assessment of screening research programme required to ascertain exact incidence in UK population and evaluate prevention of neurological disease in screened and treated individuals. Requires tandem MS</td>
</tr>
<tr>
<td>3-hydroxy-3-methyl-glutaric aciduria</td>
<td>(3) Fairly low incidence but simple treatment with prevention of neurological damage makes this a potential candidate for newborn screening. Incidence must be considered collectively as above, requires tandem MS</td>
</tr>
<tr>
<td>Biotinidase deficiency*</td>
<td>(2) Preventable severe neurological disorder, thus a candidate for newborn screening despite estimated low incidence. Screening research programme required to assess exact incidence in UK, long-term outcomes and cost-effectiveness</td>
</tr>
<tr>
<td>Congenital adrenal hyperplasia (CAH)*</td>
<td>(2) To fully meet all criteria may need screening at an earlier age (e.g. day 3) to allow availability of results by 10–12 days of age</td>
</tr>
<tr>
<td>Familial hypercholesterolaemia (FH)</td>
<td>(4) No suitable validated neonatal screening test. Benefit of intervention in neonatal period not determined</td>
</tr>
<tr>
<td>Galactosaemia</td>
<td>(4) Therapy does not prevent neurological or ovarian toxicity; long-term outcome does not benefit from early intervention through neonatal screening</td>
</tr>
<tr>
<td>Wilson disease</td>
<td>(3) Low incidence. No suitable validated neonatal screening test</td>
</tr>
</tbody>
</table>

(1) Evidence that all criteria are met.
(2) Evidence that all criteria except known natural history of disorder are met.
(3) Evidence that most criteria are met, but lack of evidence for some criteria.
(4) Evidence that some criteria are not met.

*Trials of neonatal screening recommended.
screening using tandem MS is lacking or unsubstantiated, and further research through large-scale, well-conducted and externally assessed trials is required before this technology can be universally introduced into newborn screening programmes.

Economic evidence
There was insufficient evidence in the literature to assess the economic value of screening for any disorder other than PKU. Twelve economic evaluations of neonatal screening for inborn errors of metabolism were identified.17-28 The literature covered a range of disorders, but only PKU was included in more than one study. There were eight cost–benefit studies of neonatal screening for PKU,17-21,23,25,27 all of which concluded that PKU screening is worth while in monetary terms alone. Not all of the papers included sample collection costs, but where they did these were outweighed by the net benefits of screening. This suggests that PKU screening alone justifies the collection of blood samples from neonates.

Discussion
The striking finding of this review was the lack of robust evaluation data. The absence of controlled trials meant reliance had to be put on observational studies. As it is difficult to quantify the rigour of this type of research, all being grade III or less,8 all studies were included if they fulfilled the content criteria. Agreement on the Wilson and Jungner6 criteria was then achieved by paired reviewers and eventually with the whole group. As there remains a subjective element because of the types of study used, underlying references have been provided so that readers can refer to the specific papers. Nevertheless, some evidence has been obtained for the appropriateness or otherwise of screening for a wide range of IEMs. There is clear evidence that all of the criteria are met for PKU screening, which provides a positive net benefit to the individual and to society. In addition, PKU screening by itself justifies the infrastructure of sample collection and testing of newborn blood spots.29

Table 5 summarizes the results and recommendations for each disorder. Only four disorders showed adequate fulfilment of the screening criteria to be considered for inclusion in expanded newborn screening programmes. These were biotinidase deficiency, CAH, MCAD deficiency and, possibly, GA1. For these disorders structured, co-ordinated and continuing evaluation, including economic analysis, will be necessary to provide evidence which would justify the long-term continuation of these programmes. A number of other disorders, including some of the more common disorders of organic acid metabolism, require further basic research before widespread trials of neonatal screening could be advocated.

Concern has also been expressed about the lack of consideration of the infrastructure required to deal with the infants identified by newborn IEM screening. In many areas of the United Kingdom, for example, there appears to be relatively poor liaison between the screening laboratories, midwives and other health-care personnel, and no provision is made for the co-ordinated follow-up or management of identified patients.30 Thus there is a need for a better infrastructure for notification and continued care, including parental counselling, of patients with IEMs identified through newborn screening. Few areas provide adequate information to parents about newborn screening or the IEMs screened for, nor do they currently require informed consent from the parents;3 issues which will need to be addressed if expanded screening programmes are introduced.

Newborn screening for MCAD and GA1 depends upon the use of tandem MS technology. Pressure to develop and introduce this technology is particularly acute in countries such as the United States where blood spots are taken within the first 24–48 hours after birth. However, the utility and application of tandem MS for prospective newborn screening has been demonstrated in only one centre (in the United States), with only a limited number of newborns screened.29 The technology therefore requires further assessment through primary research before it can be considered for universal introduction. The size of screening laboratories varies widely; for tandem MS technologies to be cost-effective each laboratory would need to perform a minimum number of tests per year, necessitating rationalization of current screening laboratories. There is a need for co-ordinated national policies for newborn screening that ensures the appropriate dissemination of new screening technologies and the expansion of the routine screening programme where there is good evidence of the efficacy of early detection and treatment. Any such changes must be accompanied by careful evaluation to ensure their effectiveness and cost-effectiveness in practice.

Acknowledgement
This work was funded by the NHS Health Technology Assessment Programme.

References


26 Holtzman NA. Pitfalls of newborn screening (with special attention to hypothyroidism): when will we ever learn? *Birth Defects: Original Article Series* 1983; 19: 111–120.


32 Smith I, Brenton DP. Hyperphenylalaninaemias. In:


Accepted on 9 March 1998