UK National Screening Committee

Newborn screening for methylmalonic acidaemia

External review against programme appraisal criteria for the UK National Screening Committee (UK NSC)

Version: 1

Bazian Ltd.

May 2015

The UK NSC advises Ministers and the NHS in all four UK countries about all aspects of screening policy. Its policies are reviewed on a 3 yearly cycle. Current policies can be found in the policy database at http://www.screening.nhs.uk/policies and the policy review process is described in detail at http://www.screening.nhs.uk/policyreview

Template v1.2, June 2010

Abbreviations List

Abbreviation	Meaning
C0	Free carnitine
C2	Acetylcarnitine (Number indicates the number of carbon atoms in acyl chain attached to carnitine, e.g. C2 indicates 2 carbon atoms in the acyl chain)
C3	Propionylcarnitine
C4	Butyrylcarnintine
C16	Hexadecanoylcarnitine
C4DC	Methylmalonyl/succinyl-carnitine (isomers)
CblA	Cobalamin deficiency type A. Form of MMA caused by mutation in the MMAA gene
CblB	Cobalamin deficiency type B. Form of MMA caused by mutation in the MMAB gene
CblC	Cobalamin deficiency type C. Form of MMA caused by mutation in the MMACHC gene
CblD	Cobalamin deficiency type D. Form of MMA caused by mutation in the MMADHC gene
Gly	Glycine
HCSD	Holocarboxylase synthetase deficiency
Met	Methionine
MCD	Multiple carboxylase deficiency (includes HCSD and biotinidase deficiency)
MMA	Methylmalonic acidaemia/aciduria
MMAA	Methylmalonic aciduria (cobalamin deficiency) cblA type gene
MMAB	Methylmalonic aciduria (cobalamin deficiency) cblB type gene
MMACHC	Methylmalonic aciduria (cobalamin deficiency) cblC type with homocysteinuria gene
MMADHC	Methylmalonic aciduria (cobalamin deficiency) cblD type with homocysteinuria gene
MUT	Methylmalonyl CoA mutase gene
PA	Propionic acidaemia/aciduria
R4S	Region 4 Stork
RC	Reviewer calculated figure (i.e. not presented in original publication)
SD	Standard deviation
MS/MS	Tandem mass spectrometry

Plain English Summary

Methylmalonic academia or aciduria (MMA) is an organic-acid oxidation disorder. It is a rare condition passed on to those affected through altered genes from both mother and father. The condition is caused by a lack of or faulty chemical that is needed to help break down larger substances needed by the body. These include certain amino acid (which are the building blocks of protein) and types of fatty substances. These disorders lead to a build-up of harmful chemicals. The build-up of chemicals can cause damage to vital parts of the body, such as the brain, heart, liver and nervous system. The severity of symptoms and the age that they develop varies.

Newborns affected by forms of the condition early in life often show symptoms that include poor feeding, vomiting, loss of appetite, weak muscle tone (hypotonia), and lack of energy (lethargy). In more severely affected newborns MMA can cause coma and death. Intellectual disability and delayed development can often develop in people who are affected by forms of the condition developing later in life.

Day to day treatment of MMA is mainly through changes to the diet. These changes are through eating less of certain types of proteins and by adding more of other substances (cobalamin and L-carnitine). For some people treatment helps prevent the harmful effects of the condition. Others can still develop serious problems.

It has been suggested that newborn screening for MMA would be beneficial. This is due to early treatment potentially preventing death and serious damage to vital parts of the body. The most recent review of MMA in 2009 recommended against screening due to many uncertainties. This review searched for evidence between 2001 and January 2015. The focus was on some of the areas in the 2009 review that required further evidence or were unmet.

Separate reports have been prepared for MMA and PA. This report covers MMA.

The review found that:

- MMA can be split into different subtypes. There is some evidence suggesting the severity of the condition can be predicted based on the different type of MMA they have. Measuring how much of a specific chemical (an enzyme) is active also helps predict how severely people with will be affected.

- the current screening test for MMA picks up other conditions that are not MMA. This would mean screening for MMA would not be possible without also screening for other conditions. These other conditions may not be suitable for screening.

- the quality of evidence from studies is not high enough to understand whether screening is of long term clinical benefit. This is a view supported by the current European guidelines.

- it is not clear from guidelines how to treat newborns with no medical symptoms that are picked up at screen. This could result in overtreatment of mild cases of the disease that are not the target of screening.

- there is a lack of evidence exploring the impact of the wider ethical, legal and social issues associated with screening. MMA is passed on through both parent's genes. Screening would also provide information relevant to the parents but the impact upon them has not been explored in literature.

Recommendation

The evidence considered suggests that the existing recommendation not to screen for Methylmalonic acidaemia and Propionic acidaemia in newborns should be retained.

Executive Summary

The Condition

Methylmalonic academia or aciduria (MMA) is an autosomal recessive disorder caused by mutations in a variety of genes. These mutations either give rise to MMA alone (called isolated MMA) or in combination with homocystinuria. The most common cause of MMA is mutation of the *MUT* gene, which is reported to cause about 60% of MMA cases.[1]

The *MUT* gene encodes the mitochondrial enzyme methylmalonyl-coenzyme A (CoA) mutase, which converts methylmalonyl-CoA to its isomer succinyl-CoA. This reaction is involved in the breakdown of certain amino acids, lipids, and cholesterol. It requires adenosylcobalamin – a derivative of cobalamin (vitamin B12) - to function normally. Several of the other mutations which cause MMA are due to disruptions in the production of adenosylcobalamin.

The age of onset, type and severity of symptoms is variable. Newborns with early onset forms of the condition often present with symptoms including: poor feeding, vomiting, loss of appetite, weak muscle tone (hypotonia), and lack of energy (lethargy).). These are signs of metabolic decompensation, which can cause coma and death. Later onset forms of the condition can lead to intellectual disability and delayed development. The condition can lead to damage to vital organs, such as; the brain, heart, liver and nervous system and can cause coma and death in severe cases. Even with treatment outcome in some cases is poor.

Treatment

Ongoing treatment of MMA is mainly through dietary restriction and supplementation. A low protein diet with the avoidance of certain types of amino acids, such as; methionine, threonine, valine, and isoleucine are required, in addition to supplementation with cobalamin and L-carnitine. Rapid treatment of acute metabolic crises when they do occur also forms an important part of management.

Screening

It has been suggested that screening all newborns for MMA would be beneficial due to early treatment potentially preventing death and reducing or preventing some of the serious complications like brain, heart, liver and nervous system damage. The initial screening test involves identifying raised propionylcarnitine (C3) on newborn dried blood spots using tandem mass spectrometry (MS/MS).

Previous/ Current UK NSC Review

The most recent UKNSC review of the policy on newborn screening for the organic acid disorders MMA and a similar condition, propionic acidaemia (PA), was conducted in 2009. This used evidence based on HTA reviews from 1997 and 2004 on screening for inborn errors of

metabolism and recommended against screening due to a number of uncertainties. Bazian Ltd were commissioned to undertake this current review of MMA and PA, which takes in literature between 2001 and January 2015 and focusses on some of the uncertainties raised in the last review and other recent UK NSC reviews of inborn errors of metabolism.

Separate reports have been prepared for MMA and PA. This report covers MMA.

Findings

The review found that: -

- Evidence based guidelines that suggest a reasonable understanding of enzymatic subtype-phenotype correlation. While some forms of MMA have been suggested to be earlier onset and/or poorer prognosis than other forms, the relationship was not clear cut, and many individuals carry private mutations, both of which limit the ability to predict outcome.

- The current screening test has a poor predictive value (usually less than 20%) and the initial screen cannot distinguish between PA and MMA due to them utilising the same markers (mainly C3 and ratios involving C3). The timing of the test is also of concern, with many babies presenting with clinical symptoms before screening results.

- No large, robust studies, allowing for comparison of treatment outcomes from screened and unscreened populations were identified. Current European guidelines support the conclusion that available data has not yet determined whether newborn screening for PA or MMA are of long term clinical benefit.

- The identified treatment guidelines did not give explicit recommendations about management of asymptomatic individuals identified through screening, or use specific genotype or level of PCC enzyme activity to guide management. This, and a lack of an accurate prediction of prognosis, could potentially result in treatment of asymptomatic, mild cases of the disease following screening that is aimed at more severe cases of metabolic decompensation.

- Despite parents of the affected newborns being, by default, carriers of the mutation, no studies were identified in the update search which explored the implications of the carrier state identified as a result of screening. Additionally, no direct evidence was identified that explored the impact of newborn bloodspot screening for MMA on wider ethical, legal, or social issues.

Given the rarity of MMA, it is likely a prospectively constructed international study or registry would be needed to gain sufficient evidence to compare the impact of treatment following screening versus treatment following clinical detection.

Recommendation

The existing recommendation not to screen for Methylmalonic acidaemia and Propionic acidaemia in newborns should be retained.

Introduction

Methylmalonic acidaemia

Methylmalonic academia or aciduria (MMA) is an autosomal recessive disorder caused by mutations in a variety of genes. These mutations either give rise to MMA alone (called isolated MMA) or in combination with homocystinuria.

The genes thus far known to give rise to MMA when mutated are listed in **Table 1**.[1, 2] The most common cause of MMA is mutation of the *MUT* gene, which is reported to cause about 60% of MMA cases.[1] There may still be other genes which cause MMA which have not as yet been identified, as causative mutations in these genes are not always identified.

The *MUT* gene encodes the mitochondrial enzyme methylmalonyl-coenzyme A (CoA) mutase, which converts methylmalonyl-CoA to its isomer succinyl-CoA. This reaction is involved in the breakdown of certain amino acids, lipids, and cholesterol. It requires adenosylcobalamin – a derivative of cobalamin (vitamin B12) - to function normally. Several of the other mutations which cause MMA are due to disruptions in the production of adenosylcobalamin. Some of the enzymes involved also convert cobalamin to methylcobalamin, which is required for the normal function of methionine synthase, which produces methionine from homocysteine – mutations affecting this pathway are the ones which give rise to homocystinuria as well as MMA.

Newborn screening for MMA is through tandem mass spectrometry (MS/MS) on newborn dried blood spots. MMA causes abnormal profiles of acylcarnitines in the blood, mainly raised propionylcarnitine (C3), which is the primary screening marker used of the condition. This marker is also raised in the related organic acid disorder propionic acidaemia (PA), as well as the neonatal form of multiple carboxylase deficiency (MCD), and maternal vitamin B12 deficiency. There are also a number of secondary markers for MMA which can be used in screening, largely ratios of C3 to other analytes.

Confirmatory tests for MMA include assessment of urinary organic acids including methylmalonate, assessment of plasma acylcarnitines and amino acids, as well as DNA analysis. MS/MS and urinary markers which can be used in the screening and confirmation of MMA are summarised in **Table 2**.

Basis for current recommendation

The most recent UKNSC review of the policy on newborn screening for the organic acid disorders MMA and PA was conducted in 2009. It used evidence based on HTA reviews from 1997 and 2004 on screening for inborn errors of metabolism.

It concluded that universal screening for these disorders should not be offered due to uncertainties over:

- UK incidence
- The timing of the screening test and subsequent effectiveness of intervention (particularly those with neonatal onset)
- Sensitivity and specificity of the screening test using MS/MS
- The ability of the screening test to pick-up milder forms of MMA

Table 1: Genes which when mutated cause MMA

Gene	Protein encoded	Condition caused
MUT	MUT protein (Methylmalonyl-CoA mutase)	Isolated MMA *MUT*^{0} mutations cause no enzyme activity *MUT*$^{-}$ mutations cause reduced enzyme activity
MMAA	MMAA protein	Isolated MMA MMA (cobalamin deficiency) cblA type
MMAB	MMAB protein (ATP:cob(I)alamin adenosyltransferase)	Isolated MMA MMA (cobalamin deficiency) cblB type
MMACHC	MMACHC protein	MMA (cobalamin deficiency) cblC type, with homocystinuria
MMADHC	MMADHC protein	MMA (cobalamin deficiency) cblD type, come with homocystinuria (HCU) – three forms: Classic – MMA and HCU Variant 1 – isolated HCU Variant 2 – isolated MMA, vitamin B responsive (previously known as CblH)
LMBRD1	LMBRD1 protein (LMBR1 domain containing 1protein)	MMA with homocystinuria, cblF type
ABCD4	ABCD4 protein (ATP-binding cassette, sub-family D, member 4)	MMA with homocystinuria, cblJ type
MCEE	MCEE protein (Methylmalonyl-CoA epimerase)	MMA caused by methylmalonyl-CoA epimerase deficiency
SUCLA2	SUCLA2 protein (Succinate-CoA ligase, ADP-forming, beta-subunit)	Mitochondrial DNA depletion syndrome 5 (encephalomyopathic with or without MMA)
SUCLG1	SUCLG1 protein (Succinate- CoA ligase, alpha subunit)	Mitochondrial DNA depletion syndrome 9 (encephalomyopathic type with MMA)
TCN2	TCN2 protein (Transcobalamin II)	Transcobalamin II deficiency*

*Methylmalonic acidaemia is a feature of this condition,[3] but it is not usually grouped with the other types of MMA

Key uncertainties highlighted by the NSC in commissioning this external review included:

- Is the epidemiology and natural history of MMA and PA understood?
- Have any studies explored the implications of the carrier state identified as a result of screening?
- Do we have a precise and validated means of testing for MMA or PA at screen?
- Does treatment following screening add additional benefit over treatment in non-screened populations?
- Are guidelines available for the treatment of individuals identified by screening for MMA and PA?

- Have any studies explored the wider ethical, legal and social impacts of screening for MMA and PA?

Current update review

The current review considers whether the volume and direction of the evidence produced since 2001 (the search date for the most recent HTA) indicates that the previous recommendation should be reconsidered. Six main criteria will be considered, with particular focus given to areas the NSC have identified as uncertain, or supported by insufficient evidence.

Separate reports have been prepared for MMA and PA; this report covers MMA.

Table 2: Markers of the presence PA and MMA in newborn MS/MS screening[4-7]

Disorder	Primary marker for screening	Secondary markers for screening	Urine markers (confirmatory testing)
MMA	↑C3	↑C3/C2 ↑C3/C16 ↑C3/C0 ↑C3/C4 ↑C3/Met **↑C4DC** ↑Glycine ↑Lysine	**↑ methylmalonate** ↑ methylcitrate ↑3-hydroxypropionate ↑ propionylglycine*
PA	↑C3	↑C3/C2 ↑C3/C16 ↑C3/C0 ↑C3/C4 ↑C3/Met ↑Glycine ↑Lysine	**↑propionic acid†** ↑3-hydroxypropionate ↑ methylcitrate ↑propionylglycine* **↑ N-tiglylglycine** **↑2-methyl-3-oxovaleric acid** **↑3-hydroxy-2-methylbutyric acid** **↑2-methyl-3-oxobutyric acid** **↑3-hydroxy-n-valeric acid** **↑3-oxo-n-valeric acid**

Markers in bold were not reported to be elevated in the other condition. †Propionic acid reported as the confirmatory test for PA in ACMG guidance[8], but hydroxypropionate reported as raised in both MMA and PA. *Propionylglycine was reported as raised in both PA and MMA in one paper [7] but reported as raised in only PA in a second.[6]

The main criteria and key questions reviewed are summarised in **Table 3**.

Table 3. Key questions for current MMA update review

Criterion	Key Questions (KQ)	# KQ Studies Included
2	1 Is the epidemiology and natural history of PPA and MMA understood? 1a) Are there known and understood genotype/ phenotype or enzyme activity/ phenotype correlations? 1b) When do clinical presentations manifest and how does this relate to the timing of the screen in the UK (and elsewhere)	8
4	2) Have any studies explored the implications of the carrier state identified as a result of screening?	0
5	3) Do we have a precise and validated means of testing for PPA or MMA at screen? 3a) Does the current screening marker give sufficient distinction between PA, MMA and other metabolic conditions tested for using acylcarnitine analysis? 3b) Have cut off values for the screen marker been agreed and established in screened populations for each condition and are the test values acceptable?	14
10	4) Does treatment following screening add additional benefit over treatment in non-screened populations? 4a) Is there any evidence that long-term treatment in screened populations adds benefit (early mortality, neurological symptoms, quality of life) over non-screened populations? 4b) Is there any evidence through outcome data that early treatment through screen adds additional benefit over those detected clinically?	6
11	5) Are guidelines available for screening for PPA and MMA? 5a) Is there an agreed set of criteria for who should receive treatment among asymptomatic cases (eg by genotype analysis or % enzyme activity) and what treatment they should receive?	1
14	6) Ethical, Legal and Social Issues (ELSI) Have any studies explored the wider impacts of screening for MMA and PA (eg reproductive decision making, diagnostic odyssey, implications for 'informed' consent when extending to additional NBS conditions etc)?	1

A systematic literature search of studies published between 2001 yielded 1811 references addressing MMA. Of these, 329 were assessed as being potentially relevant to the key questions outlined in Table 1. These studies were further filtered at title and abstract level, and 122 were selected for appraisal at full text (some for background reading only). Additionally, the full texts

appraised for the PA review were also considered for inclusion (there was overlap in the full texts assessed). Relevant references identified in the preparation of reviews on fatty acid and amino acid disorders were also included. Broadly, systematic reviews, RCTs, other prospective studies and screening programme evaluations were prioritised. Each section below provides additional information on the evidence selection process for the given criterion.

Appraisal against UK NSC Criteria

These criteria are available online at http://www.screening.nhs.uk/criteria.

2. The epidemiology and natural history of the condition, including development from latent to declared disease, should be adequately understood and there should be a detectable risk factor, disease marker, latent period or early symptomatic stage

Current UKNSC key question

1 Is the epidemiology and natural history of MMA understood?

1a) Are there known and understood genotype/ phenotype or enzyme activity/ phenotype correlations?

1b) When do clinical presentations manifest and how does this relate to the timing of the screen in the UK (and elsewhere)?

For question (1a) any study reporting on genotype- (or enzyme activity-) phenotype correlations was included. Studies where the abstract (or full text) only reported on mutation spectrum without indicating that relationship with phenotype was assessed were excluded. Only studies looking at the correlation between genotype or enzyme activity and clinical phenotype (i.e. age at onset, symptoms, clinical outcome) were included. Studies assessing only non-clinical (intermediate) phenotypes e.g. mRNA levels, or metabolite levels were excluded.

For question (1b) studies reporting on the age at onset of MMA symptoms in non-screened populations were included. Only studies which reported the median age at onset, or indicated how age at onset was spread across the individual weeks in the month after birth and afterwards were included. Studies presenting less detailed information about early onset cases e.g. just as 'within 1 month of birth' or 'within 1 year of birth' were excluded. This is to allow better comparison of age at onset with the age at which newborn screening and resulting diagnosis would place, as this would be within the first few days and weeks of life. Studies of screening included under criterion 5 which indicated the percentage of screen-identified cases that were symptomatic at time of diagnosis following screening were also included.

Studies with fewer than 30 participants were excluded from both KQ1a and 1b, as small studies are less likely to be able to convincingly illustrate correlations between genotype and phenotype, or typical age at onset, particularly across the different enzymatic types of MMA.

Description of the evidence

Overall, 42 studies were assessed as potentially relevant for this key question during title and abstract sifting and further assessed at full text. In addition, 5 studies extracted for criterion 5 were also relevant and described here. The results for the sub-questions are reported separately below.

1a) Are there known and understood genotype/ phenotype or enzyme activity/ phenotype correlations?

Six case series were identified which included more than 30 MMA patients (range 32 to 273), [9-14] and one European guideline was also of relevance.[6] These seven studies were included and assessed.

Results

Isolated MMA

Three of the case series (n=273, n=82, n=32) and a European guideline reported on the correlation between the mut^0, mut^-, cblA and cblB forms of isolated MMA and phenotype. [6, 9, 10, 14] They did not assess the correlation between individual mutations and clinical outcome. The types were identified by studies in fibroblasts, and genetic defects were either not reported [10] or reported to be undertaken only rarely.[9, 14]

The guideline concluded that the clinical phenotype of isolated MMA was influenced by the underlying enzymatic defect (mut^0, mut^-, cblA, cblB, and cblD-variant 2), and genotype (*MUT*, *MMAA*, *MMAB*, *MMADHC*), based on case series, case reports, and expert opinion.[6] It stated that residual enzyme activity (not explicitly stated, but likely to refer to residual mut activity in mut^-) and vitamin B12 responsiveness (which is more common in some enzymatic types than others) can be associated with better outcome.

The included case series generally supported a correlation between an enzymatic type of MMA and clinical phenotype, with mut^0 and cblB having more severe or earlier onset phenotypes than mut^- or cblA. However, the number of cblB cases was relatively small, so may not be representative of the full clinical spectrum.

The largest case series (n=273) found that patients with mut^0 MMA had significantly worse survival than patients in the other enzymatic subgroups (p=0.008), but that this difference did not remain after adjusting for age at onset of symptoms, which was the strongest predictor of survival.[9] A second case series (n=83) reported a trend towards earlier onset of symptoms with mut^0 MMA (p=0.06).[10] If enzymatic subtype influences age at onset, which is itself linked to survival, adjusting for it would potentially remove an effect of enzymatic subtype on survival, even if one existed.

In the largest case series, those with mut^0 MMA were also more likely to have motor handicap (p=0.013), but enzyme subgroup was not significantly linked to number of metabolic crises or chronic renal failure. Cobalamin responsiveness was not found in any mut^0 patients in the two larger case series, with other groups showing varying proportions of responders. [9, 10]

The smaller case series (n=32) generally supported findings of the other two, but did not present any statistical comparisons between groups.[14]

CblC

Three case series (n=118, n=41, n=37) reported on genotype-phenotype correlations in cblC MMA specifically.[11-13]

In general, they found some correlation between specific mutations in the *MMACHC* gene and age at onset, with certain mutations (c.271dupA and c.331C>T) mainly found in individuals with early onset disease and the c.394C>T mutation tending to be associated with later onset. However, even homozygotes for these mutations showed a range of ages at onset, suggesting other influences on age at onset. Many cblC mutations are private, and many individuals are compound heterozygotes, complicating genotype-phenotype analysis.[11]

1b) When do clinical presentations manifest and how does this relate to the timing of the screen in the UK (and elsewhere)?

Results

In the blood samples for UK Newborn Screening Programme Centre guidelines state that newborn blood spot screening should be collected on day 5, although in exceptional circumstances can be collected up to day 8. These guidelines also state that parents should receive screening results by the time the baby is 6 to 8 weeks old.[15] Practically, this time period applies largely to screen negative results, and there are condition specific standards for referral to specialist services following a screen positive result.[16] The standards from 2013 state, for example, that all babies in whom phenylketonuria, congenital hypothyroidism (on first sample) or medium-chain acyl-CoA dehydrogenase deficiency (MCADD) is suspected should attend their first clinical appointment by 14-17 days of age (these conditions have the earliest referral age).[16]

Age at onset or diagnosis

Six case series gave information on age at onset or diagnosis.[9-12, 14, 17] **Table 4** summarises their results.

Table 4: Age at onset and diagnosis of cases presenting symptomatically

Study, country	N and enzymatic type breakdown	Age at onset (AAO) or diagnosis (AAD)
Horster et al. 2009[9]	N=239 mut^0, mut^-, cblA/B*	*AAO:* Median 2 weeks; range reported as 1st week of life to 4.5 years, with 3 outliers up to 20 years 54.8% had AAO within 4 weeks of birth *AAD:* Median 7 weeks; range 1 week to 4.5 years – may exclude outliers
Horster et al. 2007[10]	N=77 41 mut^0, 8 mut^-, 18 cblA, 10 cblB	*AAO:* Overall median NR Median by type (ranges displayed graphically): • mut^0: 5 days • mut^-: 75 days • cblA: 25 days • cblB: 10 days *AAD*: NR
Merinero et al. 2008[14]	N=32 14 mut^0, 5 mut^-, 9 cblA, 4 cblB	*AAO:* Overall median NR Ranges by type (medians NR; AAO not available for all cases): • mut^0: 2 days to 1.5 months • mut^-: 4 days to 2 years • cblA: 2 to 8 months • cblB: 1 to 6 days *AAD:* Overall median NR Ranges by type (medians NR; AAD not available for all cases): • mut^0: 3 days to 9 months • mut^-: 1 to 2 years • cblA: 7 months to 8 years • cblB: 1 to 6 days

Study, country	N and enzymatic type breakdown	Age at onset (AAO) or diagnosis (AAD)
Karam et al. 2013[17]	N=34 23 isolated MMA (not further specified, 5 identified through newborn screening), 11 cblC MMA	*AAO:* NR *AAD:* Overall median NR, median by type: • isolated MMA: median 1.4 years; range 6 days to 14 years • cblC: median 3 months; range 1 month to 12 years
Lerner Ellis et al. 2009[11]	N=118 All CblC	*AAO:* Median NR, range 1 day to 14.4 years: • 15.3% within the first week of life • 5.9% >1 to 2 weeks • 13.6% >2 weeks to 1 month • 5.9% "neonatally" (not further defined, usually refers to first 4 weeks of life) • 25.4% >1 month to 1 year or in "infancy" (not further defined) • 12.7% after 1 year • Remainder (20.2%) did not have AAO recorded, or not presenting clinically *AAD*: NR
Nogueira et al. 2008[12]	N=41 All CblC	*AAO:* Median 3 months; range 6 days to 25 years. *AAD:* Median NR; range 2 days to 25 years (unclear why lower figure outside of range of AAO). • 4.9% within 1 week of birth • 22.0% >1 week to 1 month • 51.2% >1 month to 1 year • 22% after 1 year

*Numbers by type not reported for symptomatic cases alone; AAO age at onset; AAD age at diagnosis; NR not reported

The largest study (n=239 cases presenting symptomatically) reported a median age at onset of symptoms in isolated MMA as 2 weeks, with the range reported as 1st week of life to 4.5 years.[9] However, this range - and possibly the median - did not take into account three "outliers" with onset at between 7 and 20 years of life. Median age at diagnosis in those presenting with symptoms in this study was 7 weeks, with a reported range of 1 week to 4.5 years.

The age at onset did appear to depend to some extent on the type of MMA, with mut^0 and cblB forms tending to become symptomatic earlier than mut^-, cblC and cblA cases, although there was overlap. The studies consistently showed a potentially wide range of age at onset, with some studies reporting cases with age at onset and age at diagnosis in adolescence or adulthood.

With the timings reported in the studies at least some affected babies would be expected to present with symptoms before the results from screening are known.

Presence of symptoms at screening

One of the case series[17] and five of the screening programme evaluation studies, discussed in KQ3 of this review, reported on the percentage of children who were symptomatic at the time of screening, receipt of screening results, or diagnosis.[18-22]

These studies included between 2 and 14 screen detected individuals with MMA (the screening programme evaluations individually screened up to around 1.5 million newborns to identify these cases). **Table 5** summarises their results. They reported that between 0% and 66.7% of these cases were symptomatic either at the time of screening, or between screening and diagnosis. In two of the studies it was clear that about a third of cases had presented clinically with metabolic decompensation before the screening result was available, suggesting that these reports were not influenced by knowledge of the result.[18, 22]

One study reported that all screen detected cases either presented clinically or had 'signs of metabolic decompensation (hyperammonemia, hypoglycaemia, or metabolic acidosis) when they were admitted to confirmatory medical centres', but it was unclear whether the latter were symptomatic.[21] A study which was likely to have an overlapping population also reported that 100% of screen detected cases had metabolic decompensation when they were admitted to confirmatory medical centres.[22] It stated that symptoms included poor activity, poor appetite, vomiting, and shortness of breath, but the proportion with these symptoms was not explicitly reported.

Table 5: Proportion of screen detected newborns with MMA with symptoms at the time of screening or diagnosis

Study	Number and enzymatic type of cases	Age at screening (AAS)/ diagnosis (AAD)	% symptomatic at screening/diagnosis
Karam et al. 2013[17]	N=5 All isolated MMA*	AAS: Unclear ('detected' by NBS at 7-10 days) AAD: Unclear	0% (at 'detection' by NBS)
Lund et al. 2012[18]	N=3 (1 cblA, 1 cblB, 1 mut^{-})	AAS: 2-9 days (results 2-7 d later) AAD: NR	33.3% (presented clinically before screening result)
Shigematsu et al. 2010[19]	N=2 (1 mut, 1 vitamin B12 responsive*)	AAS: 5-8 days (age at screening result NR) AAD: NR	50% (had symptoms at screening)
Schulze et al. 2003[20]	N=3 (2 cblC/D, 1 'variant' isolated MMA*)	AAS: 3-7 days (age at screening result NR) AAD: NR	66.7% (symptomatic before screening result)
Niu et al. 2010†[21]	N=14 (13 mut^{0}, 1 cblC)	AAS: 1-3 days (age at screening result NR) AAD: NR	30.8% (presented clinically before screening result) Remainder all had 'signs of metabolic decompensation' on admission for confirmatory testing (i.e. before diagnosis) but whether symptomatic was unclear
Cheng et al. 2010†[22]	N=7 (all mut^{0})	AAS: 1-3 days (age at screening result NR) AAD: NR	100% had metabolic decompensation on admission for confirmatory testing (i.e. before diagnosis), at least some were symptomatic

*Not further defined; †Likely to be overlap of cases; AAS age at screening; AAD age at diagnosis; NR not reported

Overall comments on the evidence for KQ1

There were few reasonably sized studies (only 2 case series with >100 participants). In addition, once participants in the case series are divided up into enzymatic subtypes this renders individual groups much smaller, particularly the less common subtypes. In the screening studies looking at the presence of symptoms in screen detected cases the number of cases in each study was very small (<10 cases in all but one study).

None of the studies explicitly covered the full spectrum of enzymatic types of MMA. There may also be overlap in the patients included in at least some of the reported studies, which would further reduce the size of the population being described.

The fact that the condition is rare limits the ability to collect and analyse large groups of individuals. To obtain large groups of patients requires multinational studies, often spanning a

long period of time, over which quality of treatment can change, and confound correlation between MMA enzymatic subtype or genotype with outcomes. This was supported by the largest case series, which found that survival improved over time, with those diagnosed in the 1990s and 2000s living longer than those diagnosed before.[9]

This study also supported the existence of differences in treatment between different centres. While age at onset may not be confounded by treatment in those clinically detected, any later outcomes could be. This large study took into account differences in follow up in its survival analysis and also analysed other potential factors which could affect outcome, which allowed an assessment of the potential for confounding. However, most studies did not carry out statistical analyses or consider the impact of factors other than enzymatic type of MMA or genotype.

In most cases how the individuals were selected was not reported. The largest study relied on information being provided by multiple European centres in response to a survey, and therefore information may not be complete.[9]

Most of the individuals in the case series describing genotype-phenotype correlations and age at onset appeared to be clinically detected. Individuals with milder manifestations of MMA may not have presented or may have presented with non-specific symptoms and not been diagnosed or died before diagnosis. This could potentially bias the spectrum of disease represented in these studies.

In the studies looking at the presence of symptoms in screen detected cases, once screening has detected individuals their current and past symptoms may be attributed to MMA and interpreted differently than they would have been without the screening result. In most cases the wording stated that the children had presented clinically or were symptomatic at screening (i.e. before screening result), rather than simply having symptoms at the time of the screening result or admission, when the infant would come under more scrutiny. In two of the studies it was not possible to determine whether the symptoms were sufficient to warrant presenting for medical care in all cases.

Summary: Criterion 2 partially met.

One evidence-based guideline and the available case series suggest a reasonable understanding of enzymatic subtype-phenotype correlation. Not all outcomes analysed in the case series were significantly linked to enzymatic subtype, and there was overlap in phenotype between different enzymatic subtype groups. Evidence quality was limited by the few studies, small study size, poor control for confounders, a lack of statistical comparison between groups and of clarity over case selection.

Three case series suggested that mut^0 and cblB forms of MMA have earlier onset and/or poorer prognosis than cblA or mut^- forms. There was also some indication from three case series that certain mutations in the cblC form of MMA are more commonly associated with early or late onset. However, the relationship was not clear cut, and many individuals carry private mutations, both of which limit the ability to predict outcome.

Evidence from six case series (n=32 to n=239) suggest that some affected babies present with symptoms in the first weeks of life (particularly those with more severe forms), before screening results would become available. This was supported by five large screening programme evaluations (which identified 2 to 14 cases) that suggested many babies identified by screening are symptomatic at the time of screening, or between screening and diagnosis.

4. If the carriers of a mutation are identified as a result of screening the natural history of people with this status should be understood, including the psychological implications.

Current UKNSC key question

2) Have any studies explored the implications of the carrier state identified as a result of screening?

For this question any studies assessing the psychological implications of carrier detection as part of or due to the screening programme (for example, through cascade testing) would have been included. In addition, studies which reported whether carriers of mutations which cause MMA were being identified by screening would also have been included.

Description of the evidence

Overall, two studies were identified as potentially relevant during title and abstract sifting and further assessed at full text. Neither of these studies assessed the psychological or other implications of being a carrier of a mutation which causes MMA.

No studies were identified which assessed levels of disease markers in blood spot screening for newborn carriers of a mutation in a gene(s) which causes MMA. None of the studies on screening programmes assessed for Criterion 5 reported that newborn carriers of MMA causing mutations were being identified by newborn blood spot screening. However, as mutation analysis is not routinely carried out on those who turn out to be false positives, it is not clear whether this group might include a higher proportion of carriers.

As the mutations which cause MMA are recessive, both parents must be carriers of the mutation (unless it is a de novo mutation occurring in the gametes). Other family members may also be carriers who could be identified by cascade testing.

Results

No relevant studies were identified.

Summary: Criterion 4 unclear if met.

No studies were identified in the update search which explored the implications of the carrier state identified as a result of screening. Newborn carriers have not been reported to be picked up by the screening programmes. However, parents of the affected newborns are by default carriers, and cascade testing could identify other carriers in the family.

5. There should be a simple, safe, precise and validated screening test.

Current UKNSC key question

3) Do we have a precise and validated means of testing for MMA at screen?

a) Does the current screening marker give sufficient distinction between MMA, PA, and other metabolic conditions tested for using acylcarnitine analysis?

b) Have cut off values for the screen marker been agreed and established in screened populations for each condition and are the test values acceptable?

Description of the evidence

Overall, 50 studies were identified as potentially relevant during title and abstract sifting and further assessed at full text. Prospective cohort studies and screening programme evaluations assessing the clinical validity of the tandem mass spectrometry screening of newborn dried blood spots test for identifying newborns with MMA were prioritised. In addition, guidelines were included where they provided background on screening recommendations and practice.

Studies were included if they provided figures to indicate true and false positive results specifically for MMA. Studies presenting only true positives without a false positive rate or cut offs were excluded, as were studies which did not present figures which allowed estimation of test performance for detection of MMA specifically. Cases of PA identified through MMA screening were treated as false positives for this report.

Studies solely reporting performance of tests on individuals presenting symptomatically, or analytical validity of tests on selected individuals known to have or not have the disorders were excluded. Due to the rarity of MMA, studies including fewer than 30,000 screened infants were excluded, as they are unlikely to yield any cases. Only one study was excluded on the basis of size.

Of the 50 studies assessed at full text, 9 studies of newborn screening programmes were included in the final analysis. The main reasons for exclusion were not providing the required figures to assess clinical validity of MMA screening, or assessing analytical rather than clinical validity.

In addition, 3 guidelines addressing screening identified in the search were included. [6-8] Two relevant papers on screening programme analyte cut-offs and diagnostic guidelines for newborns who screen positive from the reviews on newborn screening for the fatty acid and amino acid disorders was also included. [5, 23]

One additional paper is cited for background information about markers used in screening.[4]

Background – guidelines addressing screening

Three US guideline on newborn screening and subsequent diagnostic follow up for conditions including PA were identified.[7, 8, 23] Two of these guidelines recommended screening for PA,[7, 8] while the third only recommended follow up procedures after such screening based on expert consensus.[23] A more recent European guideline from 2014 did not recommend for or against newborn screening.[6]

In 2006 the American College for Medical Genetics published a practice guideline on newborn screening based on expert opinion and a systematic review of research evidence.[8] It recommended that MMA caused by MUT, Cbl A, B, C and D as core conditions which should be included in a uniform newborn screening panel. Other forms of MMA were not specifically considered, possibly due to the underlying genetic defects not being characterised at that time.

On the basis of its guidance the ACMG generated and keeps updated ACTion (ACT) sheets that describe the short term actions for a health professional if an infant screens positive, and algorithms on the screening and diagnosis of the conditions.[3, 24] **Table 6** summarises their recommended diagnostic evaluation of infants screening positive for raised C3 levels, and guidance on the interpretation of results. These documents identify causes of MMA not listed in the original ACMG recommendations (SUCLA2 deficiency, transcobalamin II deficiency, and Cbl F).

Table 6: ACMG recommendations for diagnostic assays in newborns with raised C3 levels identified in newborn bloodspot screening

• Diagnostic evaluation of newborns with elevated C3 includes assay of urine organic acids (UOA), plasma acylcarnitines (PACYL), plasma homocysteine		
Assay results	**Condition indicated**	**Optional confirmatory tests**
Plasma C3 high Methylmalonic acid in urine Plasma homocysteine normal	Methylmalonyl-CoA mutase (Mut^-, Mut^0) deficiency, or CblA deficiency, or CblB deficiency	Mut enzyme assay Cbl complement studies in fibroblasts
Plasma C3 high Methylmalonic acid in urine Plasma homocysteine high	CblC deficiency, or CblD deficiency, or CblF deficiency, or Transcobalamin II, or Vitamin B12 deficiency	Cbl complement studies in fibroblasts
Plasma C3 and C4DC high Methylmalonic acid in urine Plasma homocysteine normal	Succinyl-CoA synthetase (SUCLA2) deficiency	*SUCLA2* DNA sequencing
Plasma C3 high Propionic acid in urine Plasma homocysteine normal	PA (propionic acidaemia)	PCC enzyme assay in fibroblasts
All values normal	False positive, consider maternal vitamin B12 deficiency	NA

The ACMG guidelines were developed in the US and relied to an extent on expert opinion as to the value of screening for the individual programmes. They were also considered in the context that newborn screening for various conditions is already being provided in some states, with exact conditions being screened for differing. This guideline may not be applicable to a UK setting, as views of experts in the UK may differ, and the context also differs, as screening is decided on nationally and is currently limited to a small panel of conditions.

The evidence available at that time was generally limited, and the two experts assessing the evidence identified on MMA differed, with one considering it as largely based on expert opinion, case reports, and reasoning from first principles, and the other considering it to be from well-designed studies in relevant populations, or only having minor limitations. The evidence itself was not presented, so it was unclear why this discrepancy existed.

For conditions with multiple forms with differing severity and age of onset, the ACMG decisions were based on the more severe and treatable forms of the conditions. Whether MMA was one of the conditions to which this statement applied was not explicitly stated. The guideline did note that the differences in natural history and treatment between the different forms of MMA was a reason for subdividing these conditions on the newborn screening panel list. These treatments were not specified in the ACMG document, and no specific recommendations on different treatments by genetic type were identified for this review (see Criterion 10).

The ACMG noted the potential for bias in the assessment of the severity of some conditions: without screening, the more severe forms are noticed first, biasing what is known about the effects of the condition. They also noted that until a large general population has been studied, understanding of the performance of the screening tests (in terms of the range of manifestations it identifies) is limited.

In 2009, the National Academy of Clinical Biochemistry published guidelines on the follow up testing of newborns identified through MS/MS screening.[7] These guidelines were prepared by experts based on an assessment of the peer-reviewed literature. It strongly recommended newborn screening for MMA (mut, Cbl A, B, C and D), as well as PA, as it judged that there was good evidence that it improves important health outcomes, and that benefits substantially outweighed the harms. It provided follow-up recommendations for individuals screening positive.

The New York Mid-Atlantic Consortium for Genetic and Newborn Screening Services similarly published consensus guidelines on the follow up testing of newborns identified through MS/MS screening in 2010,[23] with the latest update of these published in 2014.[25] These guidelines were based on expert consensus only, and not explicitly on the research evidence.

These US guidelines may not be applicable to the UK context, particularly as they incorporated expert opinion and the UK viewpoint may differ. Other countries may also have similar national guidelines as to whether PA screening is offered and subsequent diagnostic and treatment approaches, but these were not identified in the search.

The European guideline published in 2014 noted that newborn screening for MMA (and PA) was feasible, but did not recommend for or against, and therefore did not provide guidelines as to how it should be carried out.[6] It did provide recommendations relating to diagnosis of individuals presenting clinically, and some of these are also relevant for those identified through screening.

As this is a recent European guideline, using Scottish guideline methodology and with some members of the guideline development group coming from the UK, it is likely to be more applicable to the UK setting than the US guidelines. The evidence available was reported to be not of a high quality, and recommendations were mainly based on case series and case reports, and expert opinion.

Evidence from newborn screening programmes

The 10 studies on newborn screening programmes are summarised in **Table 7**, with additional information in the Appendix. Studies ranged in size from 146,000 (Japan, not all included in performance calculations)[19] to 1,321,123 (Taiwan).[21]

DNA testing was not reported as being carried out to confirm or rule out diagnosis in all of the studies. Therefore true mutational status of those described as true or false positives, or true or false negatives was not always known. The lack of systematic follow up in the studies to identify false negatives means that it is likely that some will have been missed. This may particularly be an issue for any cases or forms of MMA with later onset or milder phenotype, who might not present with metabolic decompensation in the newborn period.

Results from the studies may be applicable to a UK setting, as the UK has a newborn MS/MS screening programme. However, there was variation in the protocols used in the different studies (for example, in the age at which the dried blood spots were taken, range of markers and cut-offs used), which may reduce applicability to the UK of some studies. The variability of the screening protocols is also likely to impact on the results and test performance, rendering direct comparisons between different studies and wider generalisability difficult.

Variations in the incidence of MMA in the UK and other countries may also impact on test performance, and so studies from countries with more genetically similar populations (e.g. European populations) will be more applicable to the purposes of this review.

Table 7: Performance of newborn blood spot screening programmes for MMA (other conditions identified considered as false positives)

Study	Country	n	Analytes assessed	True positives† (MMA)	False positives*†	True negatives	False negatives	Specificity for MMA†	Sensitivity for MMA†	PPV for MMA†
Niu et al. 2010[21]	Taiwan	1,321,123	C3 C3/C2	14 total (13 *MUT*^{0}, 1 CblC)	61 (includes 2 PA)	1,321,048	None reported	99.995%	100%	18.7%
Weisfeld-Adams et al. 2010[26]	USA	1,006,298	C3 C3/C2 Met C4DC	26 (10 CblC, 12 mut, 4 not specified)	156 (includes 5 PA)	1,006,143	None reported	99.987%	100%	16.8%
Frazier et al. 2006[27]	USA	1997-2005 (overall): 944,078 2003-2004: 239,415	C3 C3/C2	Overall: 10 (3 *MUT*^{0}, 5 CblC, 2 not specified) 2003-2004: 5 (types not specified)	Overall: NR 2003-2004: 4 (includes 1 PA)	NR	Overall: 1 (CblA) 2003-2004: NR	NR	Overall: 90.9% 2003-2004: NR	Overall: NR 2003-2004: 55.56%
Lund et al. 2012[18]	Denmark, the Faroe Islands and Greenland	504,049	C3 C3/C2 C4DC	3 (1 mut^{-}, 1 CblA, 1 CblB)	51 (includes 2 PA)	503,992	3 (2 *MCEE*, 1 unknown cause)	99.990%	50%	5.6%
Schulze et al. 2003[20]	Germany	250,000	C3 C3/C0 C3/C2 C4DC	4 (types not specified)	206	249,790	0	99.918%	100%	1.9%

Study	Country	n	Analytes assessed	True positives† (MMA)	False positives*†	True negatives	False negatives	Specificity for MMA†	Sensitivity for MMA†	PPV for MMA†
La Marca et al. 2008[28]	Italy	160,000	C3 C4DC Gly C3/C0 C3/C4 C3/C16 MMA 3-OH-propionic acid	2 (1 mut, 1 Cbl C,D)	NR	NR	1 (CblC)	NR	66.7%	NR
Lim et al. 2014[29]	Singapore	112,289	C3 C3/C2 C3/C16 C3/Met Methyl-malonic acid	3 (1 mut/CblA, B; 2 CblC, D)	16	112,270	0	99.986%	100%	15.8%
Shigematsu et al. 2010[19]	Japan	65,000	C3 C3/C2 Methyl-malonic acid	2 (1 mut, 1 unspecified)	14 (includes 2 PA)	64,984	0	99.978%	100%	12.5%

*Cases of MMA were counted as false positives for this report; † reviewer calculated data; NR not reported (and not calculable); PPV positive predictive value

Results

3a) Distinction between MMA, PA, and other conditions tested for using acylcarnitine analysis

The primary marker for MMA in newborn blood spot screening is an elevated level of the short chain acylcarnitine propionylcarnitine (referred to as C3, indicating the number of carbon atoms in the acyl part of the molecule). All of the included studies used C3 as a marker for MMA.

Other conditions can also lead to elevated C3 levels, and data on distribution of C3 values on newborn screening from the international Region 4 Stork (R4S) data collection project shows that there is overlap between the C3 distributions seen with MMA, PA, Multiple Carboxylase Deficiency (MCD), and maternal B12 deficiency (see **Table 5**).[5] C3 is only a secondary marker for neonatal onset MCD (also known as holocarboxylase synthetase deficiency), and its primary marker (C5-OH) is not shared with MMA and PA.[4] The included studies did not report any cases of MCD being identified through MMA screening. One newborn screening report (n=160,000) reported one case of maternal MMA (CblC) being identified through newborn MMA screening.[28]

Table 5 Distribution of C3 values in MMA, normal and other populations in newborn screening from the R4S project

Population	N	Percentile (µM)		
		1st	50th	99th
All newborns	93 members*	0.57	1.75	4.74
MMA: MUT/Cbl A, B	328 cases	1.1	9.5	40
MMA: Cbl C, D	124 cases	2.8	8.4	20
PA	201 cases	2.3	14	51
MCD	15 cases	1.6	4.6	16
Maternal B12 deficiency	47 cases	*2.55	7.8	22

1 µM = 1 µmol/L; *R4S members/centres who have submitted percentile values, overall estimated to be data from 25-30 million newborns in the database. Information from McHugh et al. 2011[5]

Secondary markers for MMA include increased ratios of C3 to free carnitine (C0) and to certain other acylcarnitines (C2, C16, C4) and methionine.[4 5] The R4S data also shows overlap in the distribution of these analytes for MMA, PA and maternal B12 deficiency (and for MCD for some analytes).[5]

Most of the studies identified used the same screening markers and cut-offs to detect possible PA and MMA.[18, 21, 26, 27]

Only 4 studies explicitly reported using a slightly different screening markers or cut-offs for PA and MMA.[19, 20, 28, 29] This was either based on one first tier marker being aimed specifically at MMA but with a number of shared first tier markers with the same cut-offs,[20] or approaches using a second tier test for methylmalonic acid aimed specifically at MMA after using shared

MMA/PA first tier markers[19, 29], or both. [28] In the cases where this second tier marker is used for MMA, cases with a high level of C3 (above that used to identify individuals for second tier testing) but without raised level of methylmalonic acid are tested for PA.

The second tier test was reported to pick up maternal vitamin B12 deficiency as well as MMA.[29]

There is reported to be an increased level of succinyl/ methylmalonylcarnitine (C4DC) in MMA specifically.[4] However, this metabolite is not regularly assessed in MS/MS. Only 3 studies reported using this marker and one discontinued its use during the screening programme.[18, 26, 28]

In terms of differentiating between different types of MMA, one study used reduced methionine levels or raised C3/methionine ratio to specifically screen for cases of MMA with homocystinuria caused by Cbl C, D or F.[26]

3b) Cut offs

Cut-offs in MS/MS are set with reference to the distributions of analyte values in unaffected individuals. The R4S database project has collected and made available international data on analyte distributions from screening around 25-30 million newborns, and its main aim is to define evidence-based cut-off target ranges for all analytes.[5] They recommend setting cut-offs for elevated markers in the interval between the cumulative 99th percentile of the normal population and the 5th percentile of the disorder with elevated values for that analyte (or the disorder with the lowest values for that analyte if it is elevated in multiple disorders). There was variation between different sites in analyte values, with mean variability of median amino acid values of 23%, and 27% for acylcarnitines.

There was variability in cut-offs in the included studies, several were basing these on the mean value in an unaffected or general population sample plus between 4 and 8 standard deviations[21, 26-28] and others based them on percentiles in the unaffected or general population population.[20, 29]

This variability may be due to the programs being set up before the start of the R4S project in 2005, or when it was relatively new. The most recent publication identified[29], described using the R4S data as a guide when revising cut-offs for their population in 2010.

After initially setting the cut-offs at the start of their programs, several reported modifying these over time to optimise sensitivity and false positive rate.[18, 21, 27-29]

All of the studies reported using C3 as a marker for MMA, and cut-offs ranged from >3.3µM[28] to ≥7µM.[21] The C3/C2 ratio was the next most common marker, and cut-offs ranged from >0.15[27] to >0.39.[20] Those with lower cut-off values often designated these "borderline" values, and required a repeat test to confirm the value before referral. In addition, all studies used more than one marker for MMA, so variations in cut-offs for individual markers may reflect different combinations and approaches being used to optimise test performance.

Test performance

Overall, screening for MMA showed a similarly high specificity across all studies, ranging from to 99.92%[20] to 99.995%.[21]

As well as modifying cut-offs to reduce false positives, some programmes also aimed to do this by introducing second tier tests for methylmalonic acid itself.[19, 28, 29]

In the study using a second tier test, the specificity and positive predictive value (PPV) was higher than if they had used the first tier test alone.[29] Studies requiring collection and testing of a second sample for those with 'borderline' values also markedly reduced the number of screen positives compared to if they had used just a single borderline value as an indication of need for referral (where figures were provided).[21, 26, 27]

In the studies identified, PPV ranged from 1.9%[20] to 55.56%.[27] The study with the highest PPV was a US study, and required testing of a second sample in cases with borderline C3 or C3/C2 values, which reduced false positives. The other two programmes which also tested a second sample did not achieve as high PPVs (18.7% and 16.8%). [21, 26]

The remaining 5 studies in which a PPV could be calculated had a PPV of less than 20%, despite some using second tier testing.[19, 29] The lowest value in this range was obtained in Germany, where screening for MMA and PA was reported to have been stopped in 2005 due to the high recall and false positive rate for C3.[30] The reasons for the low PPV in this study was not clear. It used C3 ratios and methylmalonyl carnitine as a screening markers as well as C3, required blood spot samples with raised analytes to be re-tested to confirm, and then requested a repeat sample for testing if confirmed, before referral for diagnosis. The decision about whether repeat testing was needed was based on judgment of an experienced metabolic specialist using a rating system based on extent of deviation from the cutoff and overall analyte profile; this approach was not reported to be taken in other studies and may have impacted results.

Few false negatives were reported, with eight studies not reporting any (i.e. 100% sensitivity). In the three studies where false negatives were reported, there were between 1 and 3 cases, giving a sensitivity of between 50% and 90.9% .[18, 27, 28] However, there is no systematic follow up to identify false negatives, and it is likely that some will have been missed. This may particularly be the case for any cases or forms of MMA with later onset or milder phenotype, who might not present with metabolic decompensation in the newborn period.

In contrast, MMA (CblA,B) was reported to be one of the conditions with the greatest number of recorded false negatives in the R4S database up to 2010, but no further details were provided (e.g. specific numbers of false negatives) in the publication that reported this.[5]

Cheng et al. 2010 (not tabulated) reported on the MS/MS analyte levels in false positive and true positives in the newborn screening programme in Taiwan.[22] The C3 levels in the 7 MMA true positives overlapped completely with those of the false positives (MMA: 7.7 to 19.18 µM vs. false positives: 8.2 to 22.6 µM; normal value <7 µM). However, there was no overlap in C3/C2 ratio between the groups (MMA: 0.55 to 1.18 µM vs. false positives: 0.12 to 0.387µM; normal values <0.25).

They calculated that if they had used the existing C3/C2 ratio cut-offs as the primary parameter for referral (>0.2 or >0.25 in the two centres) this would have given 100% sensitivity and a PPV of 70%. If they used what was reported as the German C3/C2 ratio cutoff of >0.39, this would have given a sensitivity of 100% and PPV of 100%. On this basis they suggested re-evaluating whether C3/C2 ratio should replace C3 as the primary marker.

However, it worth noting that the German newborn screening programme has stopped screening for MMA/PA on the basis of the high recall rate with elevated C3 alone.[30] They also assessed the C3/C2 ratio in their programme, but did not decide to use this as opposed to C3 alone for detecting MMA. Other than the Taiwanese studies, 5 other studies also used C3/C2 ratio as a marker for MMA, and none of these had replaced C3 with this ratio.[18-20, 26, 27] In

addition, the R4S project data does show overlap between C3/C2 ratios for MMA, PA, MCD and maternal B12 deficiency.[5]

While some studies reported algorithms for how primary and secondary markers were combined to determine which babies were counted as screen positives,[19, 21, 26, 29] others did not state how this was reconciled (e.g. what happened if the primary marker was not raised but a secondary marker was, and vice versa).

Summary: Criterion 5 not met.

Screening for MMA and PA utilises the same markers, and these can also pick up maternal B12 deficiency and maternal MMA. It is not possible to definitively distinguish between these conditions at the screening stage.

The combinations of markers and cut-offs used in screening programmes for MMA vary. There is now an international collaboration set up to provide evidence-based cut-offs and this may lead to greater standardisation.

Specificity of screening is high (over 99.90%) but positive predictive value was usually less than 20%). In the one US study with a higher PPV (55.6%) there was no obvious differences in the screening protocol which could explain this difference.

Sensitivity is generally high, but the lack of systematic monitoring means that some false negatives are likely to be missed, particularly those with milder symptoms or later onset. MMA was reported to be one of the conditions with higher representation among the false negatives recorded in the international R4S screening data registry.

10. There should be an effective treatment or intervention for patients identified through early detection, with evidence of early treatment leading to better outcomes than late treatment.

Current UKNSC key question

4) Does treatment following screening add additional benefit over treatment in non-screened populations?

4a) Is there any evidence that long-term treatment in screened populations adds benefit (early mortality, neurological symptoms, quality of life) over non-screened populations?

4b) Is there any evidence through outcome data that early treatment through screen adds additional benefit over those detected clinically?

Description of the evidence

Overall, 27 studies were identified as potentially relevant during title and abstract sifting and further assessed at full text. Systematic reviews and randomised controlled trials (RCTs) were prioritised. As no studies of this type were identified non-randomised comparisons of outcomes of screen-detected and non-screen detected cases, or untreated cases were included. Studies

were excluded if they reported outcomes of MMA pooled with other conditions, with no data available for MMA alone.

Of the 21 studies assessed at full text, 4 were included in the final analysis,[31-34] plus one guideline for context.[6] The main reasons for exclusion were that the studies did not provide a relevant comparison of screened and clinically detected or untreated MMA cases, or did not provide separate results for MMA alone.

In addition, two studies provided indirect evidence relating to the potential impact of screening.[9, 10] They compared individuals diagnosed pre-symptomatically or in the newborn period (not necessarily through screening), with those having later or symptomatic diagnosis. These studies are also included here.

Due to the small number of relevant studies, identified results have not been split into short term and longer term outcomes. Evidence has been split into direct and indirect evidence.

Results

4) Does treatment following screening add additional benefit over treatment in non-screened populations?

Direct evidence

One European guideline developed using Scottish Intercollegiate Guideline Network (SIGN) methodology (utilising systematic review of the evidence supplemented by expert opinion as needed) noted in its recommendations that while newborn screening for MMA (and PA) was technically feasible, available data had not determined whether it was of long term clinical benefit.[6]

The 4 included studies analysed small numbers of MMA cases, ranging from 3 to 19 cases. [31-34] In three studies the clinically detected group was detected at least partially before screening was introduced.[31, 32, 34] These comparisons may be confounded by differences over time in diagnosis and management. None of the studies provided detailed information about treatments received to allow assessment of potential for confounding.

Historical control groups of children diagnosed clinically before screening was introduced are also likely to have been followed up for longer – giving a longer period in which adverse outcomes could occur. Only 3 studies attempted to take into account children's ages in their comparisons.[32-34]

Only one study reported being prospective,[33] and the remainder appeared retrospective, at least for the clinically detected cases. It was unclear whether the same follow up protocols and assessments had been used for both groups in these studies.

One of these studies found that the 2 clinically identified MMA cases in one study had more episodes of metabolic decompensation in the first and second years of life than the 1 screen-detected case, but the small numbers and use of a historical control group preclude strong conclusions.[34] The second study found that 1 out of 2 MMA cases in both the screened and unscreened groups had mental retardation before the age of three; again, small numbers limit conclusions being drawn.[33]

The third study found one death by age 6 in both the screen-detected (1/3, 33.3%) and clinically detected groups (1/16, 6.25%).[32] However, the one death in the screening group was in a child

missed by screening (a false negative) – so the outcome does not represent what would be possible with treatment.

No clearly comparable outcomes were presented for the two groups in the remaining study.[31] It was unclear whether there had been comparable assessment and follow up of the groups, and results were not clearly reported for the screened and un-screened groups.

If some cases of MMA have milder course, or are asymptomatic, the clinically detected cases may represent a group with an inherently poorer prognosis than those identified through screening. This also makes comparisons between screened and unscreened groups difficult. Studies which compare large, contemporaneous screened and unscreened populations with the same treatment protocols once cases are identified, and assess each group to compare detected incidence and type of MMA, mutations carried as well as outcome are needed. A clear genotype (or intermediate phenotype)-clinical phenotype relationship which could stratify prognosis would help to ensure the groups were comparable in terms of expected prognosis.

Indirect evidence

One case series compared those diagnosed pre-symptomatically (median age 1 week, range prenatally to 1.5 years), with those diagnosed after the onset of symptoms (median 7 weeks, range 1 week to 4.5 years, range may exclude outliers with late diagnosis).[9] The exact number of cases considered in their analysis was not explicitly reported, but overall 34 cases in the study were reported as being diagnosed pre-symptomatically (9 through screening and 25 through having an affected sibling), and 239 presented symptomatically.

Patients diagnosed pre-symptomatically were less likely to have developmental delay (p=0.03), and have less frequency and severity of motor handicap (p=0.031). There was a non-significant trend for better motor outcome with pre-symptomatic diagnosis in the mut^0 subgroup (p=0.064).

In the second case series the authors compared individuals with the CblA form of MMA who were identified and treated from the newborn period as a result of having affected siblings (n=3) and their siblings (n=3).[10] The siblings diagnosed in the newborn period had higher IQs than their affected siblings who had presented symptomatically (107 vs. 78, 117 vs. 73, and 120 vs. 106), but no statistical comparisons were provided. This comparison was only reported in the discussion of the paper. It was unclear how these individuals were selected, and it was not reported whether the newborns were showing symptoms.

There may also be confounding due to the period in which individuals were diagnosed – treatment has improved over time and those identified pre-symptomatically either through screening or having an affected sibling are more likely to have been diagnosed more recently (due to development of screening and genetically diagnosis capability over time) than those identified symptomatically.

Summary: Criterion 10 not met.

This criterion has not been met as no large, robust studies were identified which allowed comparison of outcomes of treatment in contemporaneous screened and unscreened populations. Indirect evidence from two studies comparing newborn or pre-symptomatic diagnosis (not necessarily through screening) with later, symptomatic diagnosis suggests that there may be benefits. However, the indirect nature, limited volume and quality (particularly of

one of the studies), and potential for confounding in the evidence makes drawing conclusions about the potential impact of screening difficult.

This lack of robust evidence on this issue was also noted by a recent European guideline, which also concluded that available data has not yet determined whether newborn screening for MMA of PA are of long term clinical benefit.

Given the rarity of MMA, it is likely a prospectively constructed international study or registry would be needed to gain sufficient evidence to compare the impact of treatment following screening versus treatment following clinical detection on this condition.

11. There should be agreed evidence based policies covering which individuals should be offered treatment and the appropriate treatment to be offered.

Current UKNSC key question

5) Are guidelines available for treatment of individuals identified through screening for MMA?

a) Is there an agreed set of criteria for who should receive treatment among asymptomatic cases (e.g. by genotype analysis or % enzyme activity) and what treatment they should receive?

Description of the evidence

For this question, clinical guidelines from national, state or professional organisations were included.

Of the 3 studies assessed for this question, only 1 met inclusion criteria – it described a proposed European guideline for management of MMA and PA.[6]

Results

The European guideline did not provide explicit guidance as to which asymptomatic individuals should receive treatment. [6] It focused on symptomatic presentation rather than screening, with many recommendations on the diagnosis and acute management of symptomatically presenting cases.

There were also recommendations for long term management of MMA. This included a recommendation that every MMA patient should have their response to vitamin B12 assessed, and that hydroxycobalamin should be used as long term treatment for responders. This implies that this would apply to even asymptomatic individuals, although this is not explicitly stated. It reports that all CblA identified thus far are vitamin B responsive, but this is less common in CblB and Mut^- patients, and rare among Mut^0 patients.

The long-term management described in this guideline includes various prophylactic interventions attempting to reduce the build-up of potentially toxic metabolites. This included a low protein diet, with amino acid supplementation as needed, plus L-carnitine supplementation, and oral antibiotics continuously or intermittently to control intestinal propionic acid producing bacteria. In addition, careful regular monitoring is recommended.

While, for example, the low protein diet is recommended to be titrated individually, none of the recommendations are explicitly stated as applying to only symptomatic individuals. The lack of

explicit recommendations otherwise, and the seriousness of, for example, metabolic decompensation should it arise, suggests that all individuals identified as having MMA would be considered for ongoing prophylactic interventions.

Summary: Criterion 11 not met.

A European guideline from 2014 did not give explicit criteria for which asymptomatic individuals should receive treatment. It recommended that all individuals with MMA should be tested for vitamin B responsiveness, and responders given long term hydroxycobalamin treatment. It also recommends various prophylactic long-term interventions, such as dietary protein restriction and prophylactic antibiotics to reduce the build-up of potentially dangerous metabolites and reduce the risk of developing complications, and regular monitoring.

It is likely that these measures would apply to asymptomatic individuals, for example some of those identified through screening, even though this not explicitly stated.

14. There should be evidence that the complete screening programme (test, diagnostic procedures, treatment/ intervention) is clinically, socially and ethically acceptable to health professionals and the public.

Current UKNSC key question

6) Ethical, Legal and Social Issues (ELSI)

Have any studies explored the wider impacts of screening for MMA (e.g. reproductive decision making, diagnostic odyssey, implications for 'informed' consent when extending to additional NBS conditions etc.)?

For this question any studies assessing the impact of screening for MMA specifically were included. Studies which looked at the impact of newborn screening more broadly were not included.

Description of the evidence

Four studies were identified as potentially relevant to this question during title and abstract sifting and further assessed at full text. However, none of these studies identified looked directly at the impact of MMA screening specifically on wider issues.

One European guideline reported that prenatal testing for MMA and PA is feasible.[6] This suggests that once one affected child is identified (whether by screening or clinical detection) prenatal diagnosis could be an option available for parents to inform reproductive decision making. However, no studies were identified in the search which assessed or compared reproductive decision making after screen or clinical detection of PA in a family.

Summary: Criterion 14 not met.

No direct evidence was identified on the wider impact of newborn bloodspot screening for MMA on wider ethical, legal, or social issues.

Conclusions

This report assesses newborn bloodspot screening for methylmalonic acidaemia (MMA) against select UK National Screening Committee (UK NSC) criteria for appraising the viability, effectiveness and appropriateness of a screening programme. It assessed key questions to determine if evidence published since 2004 supports a recommendation for newborn bloodspot screening for MMA in the UK.

There was some evidence of correlation between MMA enzymatic subtype and some aspects of phenotype, with mut^0 and cblB forms of MMA having earlier onset and/or poorer prognosis than cblA or mut^- forms. However, there is overlap in subtype phenotypes. Ability to find genotype-phenotype correlation is limited by high levels of heterogeneity in causative mutations.

There is evidence of varied combinations of markers and cut-offs used in screening programmes for MMA, and some variation in performance. Evidence from case series and screening programme evaluations suggest that many babies identified by screening are symptomatic by the time of screening or diagnosis.

No large, robust studies were identified which allowed comparison of outcomes of treatment in contemporaneous screened and unscreened populations. Indirect evidence suggests that there may be benefits. Given the rarity of MMA, it is likely a prospectively constructed international study or registry would be needed to gain sufficient evidence to compare the impact of treatment following screening versus treatment following clinical detection.

The volume, quality and direction of evidence published since 2001 does not indicate that the evidence has changed sufficiently to warrant recommending newborn bloodspot screening for MMA in the UK. Uncertainties remain across key criteria, including:

- Lack of a screening marker which can distinguish between MMA, PA and other conditions. The same primary marker (C3) is used in MMA and PPA screening, and is also raised in the neonatal form of multiple carboxylase deficiency and maternal vitamin B12 deficiency. In countries which screen for MMA, there is not an agreed combination of markers and cut-offs for newborn bloodspot screening programmes. Although specificity is high, these programmes have only low to moderate positive predictive values for MMA detection.
- Lack of studies on the implications of the carrier state identified as a result of screening. While newborn screening does not pick up heterozygous carriers of MMA-causing mutations, parents of affected newborns will themselves be carriers, as may other members of the families who could be identified by cascade testing.
- Despite some evidence of MMA subtype-phenotype correlation, lack of sufficient clarity on this to be able to predict prognosis and identify individuals who will have later onset or less severe manifestations of the condition, and on which to decide appropriate treatment.
- Lack of agreed criteria to decide who should receive treatment among asymptomatic cases and what treatment they should receive. All cases identified by screening are therefore likely to be treated similarly, despite differing clinical courses.
- Lack of robust evidence of a benefit of screen detection as opposed to clinical detection of MMA. Studies suggest a high proportion of MMA cases become symptomatic within the first week of life, and may therefore present before the screening results become available. No large, robust studies were identified which allowed comparison of treatment outcomes from

screened and clinically detected populations. In order to establish the additional benefit of early treatment opportunities presented by screen detection, sufficiently large studies that assess variation in outcomes according to age of treatment initiation are necessary.

- Lack of direct evidence on the wider impact of newborn bloodspot screening for PA on wider ethical, legal, or social issues.

Methodology

The draft update report was prepared by Bazian Ltd., and then adapted in line with comments from the National Screening Committee.

Search strategy

Combined searches for both MMA and PA were performed, and the search strategies are given below.

MEDLINE

1 methylmalonic acidemia.ti,ab. (392)
2 methylmalonic acidaemia.ti,ab. (105)
3 methylmalonic aciduria.ti,ab. (449)
4 methylmalonicaciduria.ti,ab. (14)
5 mcm deficien$.ti,ab. (11)
6 Amino Acid Metabolism, Inborn Errors/ (5059)
7 or/1-6 (5616)
8 Propionic Acidemia/ (89)
9 Methylmalonyl-CoA Decarboxylase/ (300)
10 propionic acidemia.ti,ab. (359)
11 propionic acidaemia.ti,ab. (129)
12 "hyperglycinemia with ketoacidosis and leukopenia".ti,ab. (1)
13 ketotic hyperglycinemia.ti,ab. (94)
14 pcc deficien$.ti,ab. (15)
15 pa deficien$.ti,ab. (46)
16 propionic aciduria.ti,ab. (38)
17 propionicacidemia.ti,ab. (12)
18 "propionyl-coa carboxylase deficien$".ti,ab. (44)
19 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 (899)
20 7 or 19 (6158)
21 limit 20 to (english language and yr="2002 - 2015") (1337)
22 (neonat$ adj2 screen$).ti,ab. (3407)
23 (newborn$ adj2 screen$).ti,ab. (5051)
24 neonatal screening/ (7379)
25 mass screening/ (82342)
26 exp Infant, Newborn/ (498630)
27 25 and 26 (4768)
28 22 or 23 or 24 or 27 (15015)
29 exp Metabolism, Inborn Errors/ (142531)
30 (inborn adj2 error$).ti,ab. (5145)
31 (metaboli$ adj error$).ti,ab. (259)
32 29 or 30 or 31 (144285)
33 28 and 32 (3036)
34 Tandem Mass Spectrometry/ (22480)

35 Spectrum Analysis, Mass/ (73663)
36 (mass adj2 spect$).ti,ab. (172907)
37 (tandem adj2 mass).ti,ab. (30548)
38 (tandem adj2 ms).ti,ab. (2118)
39 34 or 35 or 36 or 37 or 38 (209941)
40 33 and 39 (541)
41 limit 40 to (english language and yr="2002 -Current") (444)
42 21 or 41 (1705)

EMBASE

Database: Embase <1996 to 2015 January 15>
Search Strategy:
---
1 methylmalonic acidemia.ti,ab. (403)
2 methylmalonic acidaemia.ti,ab. (69)
3 methylmalonic aciduria.ti,ab. (433)
4 methylmalonicaciduria.ti,ab. (2)
5 mcm deficien$.ti,ab. (6)
6 methylmalonic acidemia/ (581)
7 or/1-6 (1089)
8 propionic acidemia/ (547)
9 methylmalonyl coenzyme A decarboxylase/ (47)
10 propionic acidemia.ti,ab. (340)
11 propionic acidaemia.ti,ab. (90)
12 "hyperglycinemia with ketoacidosis and leukopenia".ti,ab. (0)
13 ketotic hyperglycinemia.ti,ab. (76)
14 pcc deficien$.ti,ab. (6)
15 pa deficien$.ti,ab. (39)
16 propionic aciduria.ti,ab. (37)
17 propionicacidemia.ti,ab. (0)
18 "propionyl-coa carboxylase deficien$".ti,ab. (12)
19 or/8-18 (749)
20 7 or 19 (1620)
21 limit 20 to (english language and yr="2002 -Current") (1296)
22 newborn screening/ (10999)
23 (neonat$ adj2 screen$).ti,ab. (3367)
24 (newborn$ adj2 screen$).ti,ab. (6364)
25 mass screening/ (28637)
26 newborn/ (221179)
27 25 and 26 (892)
28 22 or 23 or 24 or 27 (13625)
29 exp "inborn error of metabolism"/ (128452)
30 (inborn adj2 error$).ti,ab. (4710)
31 (metaboli$ adj error$).ti,ab. (132)
32 29 or 30 or 31 (130184)
33 28 and 32 (3636)
34 tandem mass spectrometry/ (38100)
35 mass spectrometry/ (145782)
36 (mass adj2 spect$).ti,ab. (172908)
37 (tandem adj2 mass).ti,ab. (33597)
38 (tandem adj2 ms).ti,ab. (2336)

39 34 or 35 or 36 or 37 or 38 (235512)
40 33 and 39 (729)
41 limit 40 to (english language and yr="2002 -Current") (627)
42 21 or 41 (1852)
43 limit 42 to exclude medline journals (113)

Cochrane library

#1 "methylmalonic acidemia":ti,ab,kw 1
#2 "methylmalonic acidaemia":ti,ab,kw 2
#3 "methylmalonic aciduria":ti,ab,kw 1
#4 methylmalonicaciduria:ti,ab,kw 0
#5 MeSH descriptor: [Amino Acid Metabolism, Inborn Errors] this term only 12
#6 mcm next deficien*:ti,ab,kw 0
#7 #1 or #2 or #3 or #4 or #5 or #6 12
#8 MeSH descriptor: [Propionic Acidemia] this term only 0
#9 MeSH descriptor: [Methylmalonyl-CoA Decarboxylase] this term only 1
#10 "propionic academia":ti,ab,kw 0
#11 "propionic acidaemia":ti,ab,kw 1
#12 "hyperglycinemia with ketoacidosis and leukopenia":ti,ab,kw 0
#13 "ketotic hyperglycinemia":ti,ab,kw 0
#14 pcc next deficien*:ti,ab,kw 0
#15 pa next deficien*:ti,ab,kw 0
#16 "propionic aciduria":ti,ab,kw 0
#17 propionicacidemia:ti,ab,kw 0
#18 "propionyl-coa carboxylase deficien*":ti,ab,kw 0
#19 #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 2
#20 #7 or #19 14
#21 (neonat* near/2 screen*):ti,ab,kw 369
#22 (newborn* near/2 screen*):ti,ab,kw 338
#23 MeSH descriptor: [Neonatal Screening] this term only 271
#24 MeSH descriptor: [Mass Screening] this term only 4296
#25 MeSH descriptor: [Infant, Newborn] explode all trees 13259
#26 #24 and #25 119
#27 #21 or #22 or #23 or #26 565
#28 MeSH descriptor: [Metabolism, Inborn Errors] explode all trees 1795
#29 (inborn near/2 error*):ti,ab,kw 173
#30 (metaboli* near error*):ti,ab,kw 172
#31 #28 or #29 or #30 1844
#32 #27 and #31 65
#33 #32 or #20 58

Quality

Several factors were assessed to determine the quality of the identified evidence, including study design and methodology, risk of bias, directness and generalisability of the evidence. Factors that were determined to be pertinent to the body of evidence identified for each criteria are outlined in the results section as well as the comment section of the Appendix tables. The overall level of evidence was assessed by considering the quantity, quality and consistency of evidence across the body of studies for each criterion reviewed.

Appendices

Appendix number	1
Relevant criteria	KQ1, KQ6
Publication details	Horster F, Garbade SF, Zwickler T, et al. Prediction of outcome in isolated methylmalonic acidurias: combined use of clinical and biochemical parameters. J Inherit Metab Dis. 2009;32(5):630-9.[9]
Study details	Cross sectional survey of 17 European metabolic centres in 8 countries
Study objectives	To assess the long term outcome in patients with isolated MMA and identify the parameters which best predict survival, and neurological and renal complications
Inclusions	A standardised questionnaire asked for patient details. These included: date of birth, sex, clinical history during and around the time of pregnancy, age and symptoms at first manifestation, confirmation of diagnosis, subtype of MMA, drug and diet treatments received, clinical outcomes including survival, renal and neurological outcomes.
Exclusions	Infants under 3 years old at the end of the study, or who died before age 3 were excluded from the analyses looking at prognostic factors for frequency of severe metabolic decompensations. Infants under 1 year old at the end of the study, or who died before that age or from major perinatal complications were excluded from the analyses looking at prognostic factors for frequency of developmental delay. Babies who died before age 1 month or with major perinatal complications 3 were excluded from the analyses looking at prognostic factors for degree of handicap. Infants under 2 years old at the end of the study were excluded from the analyses looking at prognostic factors for onset of chronic renal failure. Non-mut^0 patients were excluded from the analyses of the effects of treatment (to avoid clinical heterogeneity).
Population	N=273 (median age 6.5 years) with confirmed diagnosis of MMA **Decade of birth:** 13 born 1960-79, 56 born 1980-89, 100 born 1990-99, and 104 2000-07 **Enzymatic type:** • 94 mut^0 • 30 mut^- • 43 Cbl A/B (7 Cbl A, 3 CblB, 33 specific enzymatic type not further assessed) • 106 enzymatic type not reported **Mutation analysis**: only done in 23 patients and identified homozygous mutations in 8 mut^0 and 1 CblA, findings in others were not reported **Vitamin B12 responsiveness:** 51 (21.9%) were responders and 182 (78.1%) non-responders; 40 did not have responsiveness reported. • mut^0 : 0% were responders, 100% non-responders • mut^- : 17.9% were responders, 82.1% non-responders • cblA/B : 67.6% were responders, 32.4% non-responders

	Ethnic background: most patients originated from Turkey (n=95), with 29 from the UK, 27 from Italy, and 27 from unspecified Asian countries, with the remainder from other countries. 49% of families were consanguineous, and 33% of patients had affected siblings
Intervention/test	NA. The prognostic factors assessed were age at onset, responsiveness to cobalamin, type of MMA, mode of and age at diagnosis, presence of symptoms at diagnosis, and decade of birth.
Comparator	Individuals with different prognostic characteristics.
Results/outcomes	**KQ1a:** The authors concluded that outcome in MMA was best predicted by enzymatic subgroup, cobalamin responsiveness, age at onset, and time period (decade) in which individuals were diagnosed. *Survival:* 29.7% (n=81) of the patients died, at a median age of 2.13 years (range 1 day to 33.9 years), this was mainly after acute metabolic crisis (n=56, 90% of those with a cause of death documented). Age at onset of symptoms had the greatest effect on survival (p<0.001), birth decade (p=0.003) and cobalamin responsiveness (p=0.042) also had strong effects (rates displayed graphically). Mortality was higher in the mut^0 group than in the other enzymatic group patients (CblA/B and mut^- group)(rates displayed graphically, p=0.008). This effect was removed if adjusted for age at onset. The most favourable outcome was for those who had late onset of symptoms (after 4 weeks of life) and were responsive to cobalamin (n=33), none of whom died during the period assessed. Those with neonatal onset (within 4 weeks of birth) born in the 90s and 00s (n=107) had better survival than those with neonatal onset born in the 70s and 80s (n=24; rates displayed graphically). This was likely to be due to improvements in treatment, neonatal care, and diagnosis over time. *Metabolic crises in first 3 years of life:* Enzymatic subgroup did not affect frequency of metabolic crises (figures not reported, p=0.656), but cobalamin responders had fewer metabolic crises (p=0.003, rates displayed graphically, median appeared to be 0 in the responders and 1 in the non-responders; crisis defined as metabolic decompensation requiring medical treatment in ICU). *Developmental delay:* Cobalamin responders were less likely to have developmental delay than non-responders, and those with neonatal onset were more likely to have developmental delay than those with late onset (results displayed graphically, significance not reported). Cobalamin responders also had less frequent motor handicap (p=0.012). Motor handicap was more common in mut^0 patients than in the other enzymatic subgroups (figures not reported, p=0.013). *Chronic renal failure (CRF):* Only the decade of birth was significantly related to risk of CRF (p<0.001). CRF was more common among cobalamin non-responders than responders

	(36% vs. 16%) and high in mut^0 patients (55%, rate in other subgroups not reported), but these differences were not statistically significant. *Growth:* There were no statistically significant differences in growth between cobalamin responders and non-responders, although most patients with retarded growth (length below 3rd centile) were non-responders (86%), as were those with microcephaly (85%; head circumference below 3rd centile). **KQ1b:** 87.5% (n=239) of patients were diagnosed as a result of symptomatic presentation (i.e. identified clinically), 3.3% (n=9) by newborn screening, and 9.2% (n=25) as a result of having an affected sibling. *Age at onset of symptoms:* median 2 weeks, range reported as 1st week of life to 4.5 years, but outliers at 7, 12, and 20 years of life. 131 developed symptoms in the first 4 weeks of life, this represented 54.8% of those presenting clinically (see comments). The most common clinical presentation was with acute metabolic crisis (68%, n=166), followed by vomiting and poor feeding (13%, n=33), failure to thrive (9%, n=23), developmental delay (6%, n=14), and other causes (4%, n=9). *Age at diagnosis*: Across all patients median 3 weeks, range reported as prenatal to 4.5 years, with outliers at 7, 13 and 14 years. In patients diagnosed before the onset of symptoms (n not reported, but overall 34 patients were pre-symptomatic at diagnosis) median 1 week (range prenatal to 1.5 years); in patients diagnosed after the onset of symptoms (n not reported, but 239 cases presented symptomatically overall, see notes) median 7 weeks (range 1 week to 4.5 years). No statistical comparison was provided. **KQ4:** Patients diagnosed pre-symptomatically were less likely to have developmental delay (p=0.03), and have less frequency and severity of motor handicap (p=0.031, severity graded as none, mild, moderate, or severe by attending physician). There was a trend for better motor outcome with pre-symptomatic diagnosis in the mut^0 group specifically (p=0.064).
Comments	The retrospective, multicentre nature of this study and the reliance on survey information meant that the data collected was incomplete, and is unlikely to have been consistently collected, recorded, and reported. However, it does mean that a large overall sample was available. **KQ1a** The metabolic centres differed in their approaches to testing individuals for cobalamin responsiveness: between 1 to 5 mg parenteral hydroxycobalamin was given and decline in urinary methylmalonic acid by either ≥30% or to <1000 mmol/mol creatinine was considered a test positive. This means the responsiveness groupings may not be homogeneous. The lack of difference in outcomes across different groups may in some cases be due to the small numbers of patients in some analyses (e.g. CRF and growth in cobalamin

	responders). Treatment was reported to differ considerably between patients and centres, and this may confound results. Most cobalamin responders received hydoxocobalamin (84%) or cyanocobalamin (8%), but 8% did not receive either; 53% of responders had a protein restricted diet, 8% had amino acid supplements, 82% had carnitine supplementation, and 8% had intermittent intestinal decontamination with antibiotics. Among cobalamin non-responders: 53% had protein restriction alone, 47% also had amino acid supplementation; 93% had carnitine supplementation; 17% had on-off intestinal decontamination, and 23% had it continuously. No significant effect of specific treatments on outcomes was found, this was reported to be as a result of treatment heterogeneity. In addition, the individuals included had varying lengths of follow up, which may impact likelihood of certain outcomes being observed. Survival was analysed using Cox regression and CRF using an accelerated failure time model to account for this. It was not clear whether the other analyses (which utilised generalised linear models, logistic regression analysis, and a proportional odds model) took this into account. The analyses did assess various possible prognostic variables, but it was unclear in most cases whether the results reported were univariate or multivariate (i.e. whether they took the other variables into account). The authors note that the lack of differentiations between Cbl A and B could be a confounder as their previous study found worse outcome for patients with CblB than in those with CblA. They suggest that this may explain the fact that enzymatic subgroup of MMA was a less important predictor of outcome in this study than the age of inset of symptoms and cobalamin responsiveness. They also note that the large number of mutations which can cause MMA (e.g. over 100 mutations in *MUT* known at that time) means that most patients have private mutations or two different mutations (compound heterozygotes) making genotype-phenotype correlation difficult to assess. The association between enzymatic subtype and survival did not remain after adjustment for age at onset. As age at onset may be related to enzymatic subtype (this was not assessed by the study), and may effectively be part of the “causal pathway” between enzymatic subtype and survival, adjusting for AAO may be an over-adjustment, which removes the relationship even though one exists. **KQ1b** It was unclear whether the age at onset of symptoms data included those identified by screening or due to an affected sibling. This could impact this outcome if these individuals are identified earlier in life and therefore monitored more closely for symptoms, or initiate treatment before the onset of symptoms thus delaying them. The proportion identified in this way was relatively small. In the assessment of age at diagnosis, it was unclear whether those diagnosed as a result of having an affected sibling all fell into the pre-symptomatic group or included some

	individuals who had symptoms by the time they were diagnosed. **KQ4** No comparisons of the characteristics of the pre-symptomatically diagnosed and symptomatically diagnosed individuals were shown. As age at onset, cobalamin responsiveness and decade of diagnosis were found to influence outcome, these (and other confounders) may be influencing outcome.

Appendix number	2
Relevant criteria	KQ1, KQ6
Publication details	Horster F, Baumgartner MR, Viardot C, et al. Long-term outcome in methylmalonic acidurias is influenced by the underlying defect (mut0, mut-, cblA, cblB). Pediatr Res. 2007;62(2):225-30.[10]
Study details	Case series of patients from 37 European hospitals
Study objectives	To assess the long term outcome in patients with isolated MMA and whether the enzymatic subgroup (mut^0, mut^-, cblA or cblB) affects this outcome.
Inclusions	Individuals with isolated MMA. Standardised questionnaires were sent out to referring physicians at diagnosis, in 1992 and in 2003. It asked about clinical presentation, major complications, anthropometric measurements, blood pressure, laboratory tests, cranial MRI, treatment, and psychomotor development. Files were also assessed by two co-authors.
Exclusions	None reported.
Population	N=83 with confirmed diagnosis of isolated MMA **Decade of birth:** 20 born 1960-79, 33 born 1980-89, 30 born 1990-97 **Enzymatic type:** Classified based on enzyme studies in cultured fibroblasts: • 42 mut^0 • 10 mut^- • 20 cbl A • 11 cbl B **Mutation analysis**: Not reported **Vitamin B (cobalamin) responsiveness:** Of the "well-documented" patients 20% (1/5) of tested cblB patients, 75% (3/4) mut^- patients, and 100% of tested cblA patients (number not reported) were cobalamin responders. Three out of another 4 mut^- patients were also responders but stopped treatment due to non-compliance; 2 did not have responsiveness assessed. None of the mut0 patients were reported as cobalamin responsive but whether they had been tested was not clear.

	Ethnic background: Most patients came from European countries, mainly Germany (n=26), Turkey (n=13), and Switzerland (n=9), with the remainder from other countries. 30% of families were known to be consanguineous, and 18 patients were siblings
Intervention/test	NA. The main prognostic factor assessed was enzymatic type of MMA; decade of birth and cobalamin responsiveness were also assessed.
Comparator	Individuals with different prognostic characteristics.
Results/outcomes	Median follow up was 18 years (range 7 to 33 years) **KQ1a:** The authors concluded that patients with mut^0 and cblB MMA had earlier onset of symptoms, more complications and deaths and more pronounced urinary excretion of methylmalonic acid than patients with mut^- and cblA MMA. *Age of onset of symptoms:* Six cases did not have this information documented. For the remainder: the mut^0 patients had the earliest onset of symptoms (median 5 days), followed by CblB (median 10 days), then CblA (median 25 days), then mut^- (median 75 days; ranges displayed graphically), the differences did not reach statistical significance (p=0.06). Of the mut^0 patients, 73% had onset in the first week of life, 60% of CblB had "neonatal onset" (not defined), as well as 55% of CblA patients, and 37% of mut^- patients. *Survival:* 37% (n=30) of the patients died, at a median age of 2.7 years (range 4 days to 22.6 years). Most deaths occurred during severe metabolic crises, but few post-mortems were carried out. There were significant differences across the groups (p=0.01 for overall comparison), with the fewest deaths in the CblA group • mut^0: 20 deaths (47.6% RC), at median age 2 years (range 4 days to 22.6 years) • Cbl B: 5 deaths (45.5% RC), at median age 2.9 years (range 7 days to 5.9 years) • mut^-: 4 deaths (40% RC), at median age 4.5 years (range 4.6 months to 8.9 years) • Cbl A: 1 death (5% RC), at age 14 days. Survival rate improved in the mut^0 group over time (p<0.005), but did not change in the other groups (p values not reported). *Anthropometrics:* Failure to thrive was most common in mut^0 patients, and least common in CblA patients (5%; statistical comparisons not provided). Median length was below age-specific means in all groups, and significantly different between the groups (p=0.02 for overall comparison). Mut^0 had the greatest reductions in length (-2.2 standard deviations [SD], range -6.2 to +0.4), with Cbl B being the next greatest (-1.7 SD, -4.9 to +1.8), and finally mut^- (-1.05 SD, range -3.1 to +0.9) and CblA (-1 SD, range -2.1 to +1) which showed similar reductions. *Chronic renal failure (CRF):* CRF differed across the enzymatic groups (p=0.005 for overall comparison), affecting 61% of people with mut^0 MMA aged over 2 years (median age at manifestation of CRF 8 years, range 2 to 18), 66% with CblB (median age 13.5 years, range 12 to 15), 21% with CblA (median age 11 years, range 6.4 to 12), and 0% of mut^-.

	Neurologic outcome: Neurologic complications were common in all groups, mainly motor disorders (30%), mental retardation (25%), and seizures (23%). No comparison between the subgroups was provided for these complications. Cognitive dysfunction was reported to differ between the groups, with mut^0 patients having the poorest outcome, but figures (for IQ) were displayed graphically, and no statistical comparisons provided. From the graph, mut0 patients had the highest rate of patients with IQ<60 or no schooling (between 25 and 30%), with about 5% of Cbl A patients showing this outcome, and none of the other groups. The mut$^-$ group had the highest proportion with IQ>90 or in normal schooling (between 80-90%), with Cbl B having the next highest (about 70%), followed by mut^0 (about 50%), and finally cblA (just under 50%). The numbers in each group in this analysis were small (24 mut^0, 5 mut$^-$, 19 CblA, 6 CblB). 7 out of 10 mut^0 patients assessed with MRI had persistent abnormalities, 5 out of 8 CblA patients, 1 out of 3 mut$^-$ patients, and 1 out of 2 CblB patients. *Gastrointestinal complications:* Two mut^0 patients had acute pancreatitis. 13 patients overall had feeding difficulties which required tube feeding, and 22 of these were mut^0. No statistical comparisons were provided for this group. **KQ1b:** 48% of patients had symptoms within the first week of life, and 52% after this (not further specified). Also see "Age at onset of symptoms" under KQ1 above. **KQ4:** In the discussion, the authors note that three individuals with CblA who were identified and treated from the newborn period as a result of having affected siblings, had higher IQs than their affected siblings (107 vs. 78, 117 vs. 73, and 120 vs. 106). No statistical comparisons were performed and no comparisons were given for other enzymatic forms.
Comments	While participants in this study were reported to be "prospectively followed" it was unclear whether this was a part of their normal clinical follow up or specifically for the purposes of the current study. Given that participants were born between 1971 and 1997 and followed up to 2004 (i.e. up to 33 years), and that questionnaires were sent to the referring physicians to collect outcome data it seems likely to have relied on routinely collected data rather than data prospectively collected for the purposes of the study. This may lead to some data being incomplete. It was unclear whether there was overlap between these participants and those in Horster et al. 2009[9], which was written by an overlapping group of authors. **KQ1a** The authors noted that there is no common consensus on evaluating cobalamin responsiveness, and that it is difficult to assess, particularly in those with low urine methylmalonic acid levels. Responsiveness was usually tested by measuring urinary methylmalonic acid after repeated administration of between 1 to 10 mg hydroxycobalamin intramuscularly or intravenously.

	Statistical comparisons were not provided for all outcomes, and in some cases where they were performed and did not reach significance (e.g. age of onset of symptoms) may be due to the small numbers of patients in some groups. Treatment was not reported and outcomes may be confounded by differences in treatment received, as well as other confounders. The analyses did not appear to be adjusted for other factors. In addition, the individuals included had varying lengths of follow up, which may impact likelihood of certain outcomes being observed. Survival was analysed using log rank tests and Kaplan-Maier curves, which should account for this. It was not clear whether the other analyses (which utilised the Kruskal-Wallis rank sum test and the Pearson's chi squared test) took this into account. The analyses did not appear to be adjusted for potential confounders such as treatment. They did consider the effects of time period of birth on survival, but it was unclear if this was adjusted for in the analyses looking at the impact of enzymatic subtype. While the authors identified a correlation between methylmalonic acid in the urine and CRF, they note that there was considerable day to day variation in individuals mainly as a result of varying protein and food intake and hydration. Nutritional parameters were not used to standardise measurements and this could introduce bias. They suggest that their findings need confirmation in international prospective multicentre follow up studies which include standardised biochemical monitoring.

Appendix number	3
Relevant criteria	KQ1
Publication details	Lerner-Ellis JP, Anastasio N, Liu J, et al. Spectrum of mutations in MMACHC, allelic expression, and evidence for genotype-phenotype correlations. Hum Mutat. 2009;30(7):1072-81.[11]
Study details	Case series of international patients from reference centres in Canada and Switzerland.
Study objectives	To assess the *MMACHC* mRNA levels in patients with different mutations associated with early and late onset CblC disease.
Inclusions	Individuals with confirmed CblC MMA diagnosed at the two international reference centres offering definitive diagnosis of CblC MMA. These centres were reported to have diagnosed the majority of known cases of CblC MMA (n=366 at that point).
Exclusions	None reported.
Population	N=118 patients had sequence and mRNA analysis in this study. MMACHC mutations were also reported for all 366 CblC patients reported to date.
Intervention/test	NA. *MMACHC* mutation.
Comparator	NA

Results/outcomes	**KQ1a:** The 118 patients had 34 different mutations, 11 of which had not been described previously. 109 patients had two different mutations, three had three mutations, three had only one mutation identified, and 3 had no causal mutations identified. Common mutations included the c.271dupA mutation (42% of pathogenic alleles), c.394C>T (20%), and c.331C>T (5%). In the total sample of 366 cblC individuals there were: • 13 nonsense mutations • 17 missense mutations • 3 possibly benign missense changes • 11 deletions resulting in frameshifts • 4 in frame deletions • 3 duplications resulting in frameshifts • 1 in frame duplication • 1 insertion • 3 splice-site mutations • 3 mutations affecting the initiation codon • 41 alleles with no mutation identified The c.271dupA and c.331C>T mutations were mainly found in individuals with early onset disease (<1 year), but not entirely. Five individuals with one c.271dupA mutation had onset at after 1 year (range 20 months to 14 years). Two individuals with c.331C>T mutations (one homozygous) had onset after 1 year (72 weeks and 4 years). *c.271dupA homozygotes:* onset from the neonatal period to 7 months. *c.331C>T homozygotes:* onset from 5 weeks to 18 months. The c.394C>T mutation was observed "frequently" in individuals with late onset disease. However, 13 individuals carrying the mutation had onset up to and including 1 year, including number of homozygotes. *c.394C>T homozygotes:* onset from 2 months to 13.5 years. The less common c.482G>A mutation had previously been reported to be associated with late onset disease (presentation after 10 to 20 years of age) but the one homozygote in this study had an onset at age 2.5 months. **KQ1b:** Overall, age at onset ranged between 1 day to 14.4 years: • 18 (15.3%) had onset within the first week of life (range 1-7 days • 7 (5.9%) had onset between >1 and 2 weeks after birth

	• 17 (13.6%) had onset between >2 weeks and 1 month after birth • 7 (5.9%) had "neonatal" onset not further defined (generally refers to first 4 weeks of life) • 30 (25.4%) had onset between >1 month and 1 year (includes "infancy" onset not further defined) • 15 (12.7%) had onset after 1 year • 5 (4.2%) were identified by newborn screening • 1 (0.8%) was identified due to having an affected sibling • 1 (0.8%) was diagnosed prenatally • 17 (14.4%) had no information recorded on onset. Two individuals had the given age noted as age at diagnosis rather than onset.
Comments	The authors note that: • clinical heterogeneity in cblC disease has been clearly demonstrated • genotype-phenotype observations are "consistent" with the clinical heterogeneity • for example, although 94 of the 96 individuals known to be homozygous for c.271dupA had presented in the first year of life, 2 presented between 1 and 4 years • non-uniform reporting of clinical histories may explain some of the variability in individuals with the same mutations • most mutations are private The individuals in the study were not related to the knowledge of the authors, but patient samples were anonymous, meaning there may have been reported more than once due to diagnosis at both reference centres. It was not reported how the 118 individual being assessed had been selected. No statistical analysis if the genotype-clinical phenotype correlation was carried out.

Appendix number	4
Relevant criteria	KQ1
Publication details	Nogueira C, Aiello C, Cerone R, et al. Spectrum of MMACHC mutations in Italian and Portuguese patients with combined methylmalonic aciduria and homocystinuria, cblC type. Mol Genet Metab. 2008;93(4):475-80.[12]
Study details	Case series of patients from Portugal and Italy.
Study objectives	To assess experience with new cblC cases from Italy and Portugal.

Inclusions	Individuals with new CblC diagnosis from Italy and Portugal, selected "after sharing and matching databases".
Exclusions	None reported.
Population	N=41 patients from Portugal and Italy.
Intervention/test	NA. MMACHC mutation.
Comparator	NA
Results/outcomes	**KQ1a:** The c.271dupA mutation was found in 55% of alleles, c.394C>T in 16%, and c.331C>T in 9%. c.271dupA homozygotes (n=14) all had early onset disease, the two c.394C>T homozygotes both had late onset disease. 60% of alleles in early onset patients were c.271dupA, 10% were c.331C>T, and 10% c.394C>T. 20% of alleles in late onset patients were c.271dupA (this included the individual diagnosed at age 25 years), 0% were c.331C>T, and 50% c.394C>T. **KQ1b:** Median age at onset was 3 months (range 6 days to 25 years). 36 (87.8%) presented in the first year of life and 5 (12.2%) after this. Age at diagnosis ranged from 2 days to 25 years. Overall age at diagnosis was [RC]: • 2 (4.9%) within 1 week of birth (2 days and 6 days) • 9 (22.0%) between >1 week and 1 month of birth (18 days to 1 month) • 21 (51.2%) between >1 month and 1 year of birth • 9 (22%) after 1 year
Comments	The number of individuals assessed was small and it was unclear how they were selected, and whether they all presented symptomatically. The study by Lerner-Ellis et al. 2009[11] included individuals from Italy and Portugal and there may be overlap. While the ages were reported as age at onset in some places in the text, they were also referred to as age at diagnosis in the table. No statistical analysis of genotype-clinical phenotype was carried out.

Appendix number	5
Relevant criteria	KQ1
Publication details	Morel CF, Lerner-Ellis JP, Rosenblatt DS. Combined methylmalonic aciduria and homocystinuria (cblC): phenotype-genotype correlations and ethnic-specific observations. Mol Genet Metab. 2006;88(4):315-21.[13]
Study details	Case series of international patients diagnosed at a reference laboratory in Canada.

Study objectives	To describe the genotype-phenotype in cblC patients.
Inclusions	Individuals with CblC diagnosis, reported in published case reports.
Exclusions	None reported.
Population	N=37 patients
Intervention/test	NA. MMACHC mutation.
Comparator	NA
Results/outcomes	**KQ1a:** 17/25 (68%) early onset patients were homozygous for c.271dupA (n=9) or c.331C>T (n=3) or carried both of these mutations (n=5). Most of the other early onset patients were reported to be compound heterozygotes for nonsense mutations or a nonsense mutation and a frameshift mutation. Two early onset patients carried one copy of the c.394C>T mutation. 4/12 late onset patients (33.3%) were homozygous for the c.394C>T mutation. They had normal development and health until adolescence or early adulthood followed by sudden neurological deterioration, except for 1 girl who was asymptomatic at age 12 and was identified after diagnosis of an older sibling. 2/12 (16.7%) late onset patients were compound heterozygotes for the c.271dupA and c.394C>T mutations, while one patient with this genotype had early onset disease undistinguishable from that of c.271dupA heterozygotes. The remaining late onset patients (6/12, 50%) carried c.271dupA and another mutation (not c.394C>T), often reported to be a missense mutation. The c.892-9_12delTTTC mutation was seen in 3 patients along with c.271dupA, with a unique presentation of haemolytic uremic syndrome. Different mutations at the same nucleotide had the potential to differ in phenotype: c.440G>A (glycine to aspartic acid) was seen in a patient with onset at 45 years who had previously been healthy, while c.440G>C (glycine to alanine) was seen in two patients with early onset and severe systemic involvement. In both cases the patient's other allele was c.271dupA.
Comments	The authors say that their results suggesting "apparent genotype-phenotype correlations" could help in the identification of mutations in cblC mutations, and that tests for the common mutations could be used in newborns identified by screening. The number of individuals assessed was small. The reference centre where these cases were one of the two contributing to the patients described in the later study by Lerner-Ellis et al. 2009[13], so they are likely to have been included in the 366 patients described, the overlap with the 118 who were specifically the focus of that study is unclear. No statistical analysis of genotype-clinical phenotype correlation was carried out. The study focused on patients with published case reports, but the basis on which these had been selected for these case reports was unclear.

Appendix number	6
Relevant criteria	KQ1
Publication details	Merinero B, Perez B, Perez-Cerda C, et al. Methylmalonic acidaemia: examination of genotype and biochemical data in 32 patients belonging to mut, cblA or cblB complementation group. J Inherit Metab Dis. 2008;31(1):55-66.[14]
Study details	Case series of international patients with isolated MMA.
Study objectives	To describe the genotype-phenotype in isolated MMA patients.
Inclusions	Patients with isolated MMA "representing the entire clinical spectrum of the disease".
Exclusions	None reported.
Population	N=32 patients from unrelated families diagnosed over 25 years. Two pairs of two patients were siblings. Cases came from Spain, Latin America, Italy, and the UK. There were: • 14 mut^0 • 5 mut^- • 9 cblA • 4 cblB
Intervention/test	NA. Genotype.
Comparator	NA
Results/outcomes	**KQ1a:** 13/14 mut^0 patients presented in the neonatal period and had a severe course of the disease, the mut^- patients had a "milder clinical and biochemical phenotype". None of these patients showed vitamin B responsiveness. 8/9 cblA patients had onset in infancy, and they had "less life-threatening episodes than the mutase-deficient patients, despite a lack of dietary compliance in some". 4/9 patients showed hydroxycobalamin responsiveness 4/4 cblB patients had clinically severe disease with neonatal onset and poor prognosis, and none of them were hydroxycobalamin responsive **KQ1b:** 16/32 patients (50%) presented in the first month of life (neonatally), 10 (31.3%) between 1 month and 1 year, 2 (6.3%) between 1 and 2 years, 1 "in infancy" (not further specified), and 3 were diagnosed due to having an affected sibling (2 prenatally). Where reported age at onset and diagnosis, and survival were as follows: • 14 mut^0 : onset at 2 days to 1.5 months; diagnosis at 3 days to 9 months; 9/14 (62.4%) died at 80 days to 12 years, those alive were aged 18 months to 18 years. • 5 mut^- : onset at 4 days to 2 years; diagnosis at 1 to 2 years; none died, those alive were aged 18 months to 19 years. • 9 cblA : onset at 2 to 8 months; diagnosis at 7 months to 8 years (where reported); none were known to have died, but 3 had been lost to follow up, at

	last follow up those alive were 3 to 21 years • 4 cblB : onset at 1 to 6 days; diagnosis at 4 days to <1 month; 3 died – 1 diagnosed prenatally was miscarried and the others at 2 and <10 days, 1 alive was aged 3 years The mut^0 and cblB patients were reported as the most clinically affected, with higher mortality rates. There was some inconsistency in individuals with the same genotype: two mut^0 patients homozygous for c.671-678dup (p.V227fs) had different clinical course – with the worse outcome in the patient diagnosed at an earlier time point (in terms of calendar year of diagnosis).
Comments	The authors concluded that factors other than genotype appear to contribute to the clinical phenotype of MMA. They also suggested that the outcome of the disease may depend more on treatment improvement over time and on number of metabolic episodes (leading to cerebral lesions) suffered by the patient at an early age than genotype. The number of individuals assessed was small, particularly in the cbl and mut$^-$ groups. It was unclear how cases were selected for inclusion. Data was missing on age at onset or diagnosis, for some cases. The differing lengths of follow up may influence e.g. mortality rates seen. There may be overlap with the European case series described by Horster et al 2009.[9]

Appendix number	7
Relevant criteria	KQ1
Publication details	Karam PE, Habbal MZ, Mikati MA, et al. Diagnostic challenges of aminoacidopathies and organic acidemias in a developing country: a twelve-year experience. Clin Biochem. 2013;46(18):1787-92.[17]
Study details	Retrospective case series of confirmed MMA patients from Lebanon.
Study objectives	To review patients diagnosed over 12 years at one centre.
Inclusions	Patients with MMA diagnosed between 1998 and 2010 after referral to the centre with a suspected inborn error of metabolism, or as a result of newborn screening (introduced 2008).
Exclusions	None reported.
Population	N=34 There were: • 23 isolated MMA (5 identified through NBS) • 11 cblC (none identified through NBS)
Intervention/test	NA. Genotype.

Comparator	NA
Results/outcomes	**KQ1b:** Median age at diagnosis: • 1.4 years isolated MMA (range 6 days to 14 years) • 3 months for cblC (range 1 month to 12 years) It was unclear if individuals diagnosed through NBS were included in the ranges given above. Newborn screening patients were reported as being detected by NBS between 7 and 10 days of age, but this was across all of the disorders being screened for, not just MMA. It was also unclear whether this "detection" referred to the timing of DBS collection or diagnosis. The cases identified through NBS were reported to be asymptomatic when detected by screening.
Comments	The study was small, and did not separate different enzymatic types of MMA. It was reported to cover all cases identified at the centre, reducing the chances of selection bias. The inclusion of cases diagnosed as a result of newborn screening will reduce age at diagnosis.

Appendix number	8
Relevant criteria	KQ1, KQ3
Publication details	Niu DM, Chien YH, Chiang CC, et al. Nationwide survey of extended newborn screening by tandem mass spectrometry in Taiwan. J Inherit Metab Dis. 2010;33(Suppl 2):S295-305.[21]
Study details	Screening programme evaluation from 3 centres in Taiwan
Study objectives	To report the incidences and outcomes of inborn errors of metabolism in nationwide newborn screening in Taiwan
Inclusions	Newborns having screening between March 2000 and June 2009
Exclusions	None reported
Population	n=1,321,123 newborns
Intervention/test	Newborn screening for MMA, PA and other conditions using tandem mass spectrometry (MS/MS) on dried blood spot collected by heel prick at 1 day after 1st feeding or 2-3 days of life. Preterm infants had a 2nd DBS taken and tested at 1 month of age. <u>Cut-off values:</u> The system used included two cut-offs: a 'positive' cutoff which warranted immediate referral, and a borderline cutoff – values between this and the positive cutoff resulted in repeat DBS sampling and testing. Borderline values initially set at mean + 4 standard deviations (SD) and positive cut-offs at

	twice borderline values. These cut-offs were modified over time to minimise false positives and negatives. C3 borderline: ≥7 µM (2 centres), ≥4.74 µM (1 centre) C3 positive: ≥12 µM (2 centres) >8.8 µM (1 centre) Or (in 2 centres) C3/C2 borderline: ≥0.2 (2 centres) C3/C2 positive: ≥0.3 (1 centre), ≥0.25 (1 centre) Borderline values needed to be confirmed in in a repeat test before an infant was referred. The timing of collection of the repeat sample was not reported. Confirmatory tests are reported below as the 'Comparator' for the screening test results as they formed the gold standard against which false and true positives and negatives were judged.
Comparator	Confirmatory tests included UOA, acyl carnitine analysis before and after 3 day carnitine loading (100 mg/kg per day; used because acylcarnitine results tended to improve on repeated samples in many OA cases), serum B12 and plasma homocysteine analysis.
Results/outcomes	**KQ3:** Requiring confirmation of borderline values in a repeat sample cut the number of cases requiring referrals dramatically (from 3,114 to 39). Results below are for the process as using confirmed borderline values and positive values to indicate referral. In addition, results are in brackets illustrate what results would have been obtained if only the positive cutoff values were utilised. True positives: 14 (13 MUT and 1 Cbl) False positives: 61 (27) (these numbers include 2 cases of PA) True negatives: 1,321,048 (1,321,082) False negatives: None reported (5) Sensitivity: 100% (64.3%) [RC] Specificity: 99.995% (99.998%) [RC] PPV: 18.7% (25%) for referred cases, 0.44% for all recalls (i.e. referrals plus those where a repeat sample was requested) [RC] <u>Incidence:</u> MMA: 1 per 94,365 [RC] PA: 1 per 660,561 [RC] **KQ1b:** <u>Symptoms among cases:</u> 4 *MUT*^{0} MMA cases were admitted to hospital due to overwhelming metabolic decompensation before screening results were available, 3 of these had severe developmental delay (age at assessment not reported; treatment initiated at 4-9 days of

	age). The other 9 *MUT*^{0} cases were showing signs of metabolic decompensation when they were admitted for confirmatory testing. 9 *MUT*^{0} cases had significant but not severe developmental delay despite early intensive treatment (started age 3-5 days, and 5/9 started before 5 days of age); the remaining *MUT*^{0} case did not have developmental delay at 10 months (started treatment at 6 days). The CblC case had mild developmental delay at age 10 months. One PA case had moderate developmental delay at 5 years, and the other none at 6 years.
Comments	No systematic approaches to identifying false negatives was described, therefore some may be missed, particularly among those with a single borderline C3 or C3/C2 value. The population described is likely to overlap with that of Cheng et al. 2010[22], which describes outcomes at two of the centres described in an overlapping time period (2002 to 2008). The cutoff values reported by Niu et al. for these 2 centres differ from those reported in Cheng et al. 2010[22]. This may be due to changes in cut-offs over time.

Appendix number	9
Relevant criteria	KQ3
Publication details	Weisfeld-Adams JD, Morrissey MA, Kirmse BM, et al. Newborn screening and early biochemical follow-up in combined methylmalonic aciduria and homocystinuria, cblC type, and utility of methionine as a secondary screening analyte. Mol Genet Metab. 2010;99(2):116-23.[26]
Study details	Screening programme evaluation from the USA
Study objectives	To report on newborn screening data from 10 patients with CblC born in New York State since 2005, including description of the impact of incorporating a new secondary marker for CblC/D/F
Inclusions	Newborns screened in New York State between January 2005 and December 2008
Exclusions	None reported
Population	N=1,006,325 newborns
Intervention/test	Newborn DBS screening for MMA (including CblC), PA and other conditions using MS/MS. <u>Cutoff values:</u> The C3 cutoff value was mean of a normal population sample plus about 8 SD (7 µM) Category 1: C3>7 µM and C3/C2 >0.2 (immediate referral) Category 2: C3>7 µM and C3/C2 <0.2 Category 3: C3 in range of 5 to7 µM and C3/C2 >0.2 For Categories 2&3, a repeat sample is requested and tested and the infant referred if C3 >7 µM , or C3 5-7 µM and C3/C2 >0.2. The timing of collection of the second sample was not reported.

	In 2005 C4DC was added as a secondary marker for MMA/PA, a value of C4DC >1 µM on initial and repeat sample leads to referral. From late 2008 (variably reported as September or November), methionine (Met) was added as a secondary marker to identify CblC, D, and F diseases. Levels of Met were considered in those falling into Category 3 before repeat sample was requested. The following approach was then taken: Category 3a: Met >13.4 µM, repeat sample requested and tested as for Category 3 above Category 3b: Met <13.4 µM, immediate referral (suspicion of CblC, D, or F) This aimed to reduce time to referral in those with a higher suspicion of being CblC, D, or F cases. Met was already being assessed in the MS/MS screen as a marker for cystathione beta synthase deficiency when elevated above a specified cutoff.
Comparator	Confirmatory tests were not explicitly reported, but DNA analysis was used, as was complementation testing.
Results/outcomes	True positives: 26 (10 MMA CblC and 12 MMA MUT identified through C3 algorithm; 4 MMA type not specified, identified through C4DC analyte; no cases of MMA Cbl D or F were identified) False positives: 156 (includes 5 PA cases, and 27 with screening results unresolved) False negatives: None reported True negatives: 1,006,143 The 27 screen positives with screening results unresolved included 2 who had died, 9 lost to follow up, and 16 'remaining open' at the time of publication. Figures have been calculated [RC] for either counting these as false positives, or excluding them from the analyses completely Sensitivity: 100% Specificity:99.98% (if unresolved screen positives considered as false positives), 99.99% (if unresolved screen positives excluded) PPV: 14.29% (if unresolved screen positives considered as false positives), 16.77% (if unresolved screen positives excluded) <u>Results by test category were:</u> Categories 1& 2: 577 requests for repeat specimens, and 144 referrals. These yielded 18 MMA [7 of which were CblC] and 5 PA cases, 100 who had false positive results, and 21 who had unresolved results (death, loss to follow up, etc.) So PPV for category1&2 for MMA was 3.1% considering first specimens only, and 12.5% for referrals [RC] (assuming unresolved results false positive) Category 3 (up to September 2008): 310 requests for repeat specimens, and 9 referrals. These yielded 2 MMA [1 CblC], 5 false positives, and 2 had unresolved results So PPV for category 3 for MMA was 0.6% considering first specimens only, and 22.2% for referrals [RC] (assuming unresolved results false positive)

	C4DC: requests for repeat specimens not reported, 30 referrals and 4 cases of MMA (0 CblC) So PPV for C4DC for MMA was 13.3% for referrals [RC] (assuming unresolved results false positive) NB The CblC cases described above only total 8, as 2 cases were diagnosed based on category 3b after this. If the new CblC/D/F category (3b) had been used, it would have detected 3 out of the 10 CblC cases. The use of the new CblC/D/F category (3b) was estimated to add an extra 3 referrals per year (based on screening results from 2008), with 2 of these being CblC cases. The new category 3b was reported to have 100% PPV for CblC, as 2 cases referred in 2008 had been confirmed as CblC, but in 2009 only 1 of 2 referred cases had CblC, bringing this down to 75% [RC]. The old category 3 was reported to have a PPV of 14.3% (1/7) for CblC, and 28.6% (2/7) for MMA or CblC up to 2008 (excluding unresolved results).
Comments	No systematic follow up of screen negatives, or monitoring for false negatives was reported, so these may have been missed. All 10 CblC cases appeared to have been detected or detectable under the original C3 algorithm, but referral for 1 case would have been quicker with the new algorithm. Low Met was noted to not be a useful screening test for CblC, D, or F in isolation as many infants have low Met in the days after birth when protein intake is low, low levels are also seen in methylene tetrahydrofolate reductase (MTHFR) deficiency

Appendix number	10
Relevant criteria	KQ3
Publication details	Frazier DM, Millington DS, McCandless SE, et al. The tandem mass spectrometry newborn screening experience in North Carolina: 1997-2005. J Inherit Metab Dis. 2006;29(1):76-85.[27]
Study details	Screening programme evaluation from the USA
Study objectives	To assess the performance of the newborn screening programme in North Carolina over 8 years
Inclusions	Newborns undergoing screening from July 1997 to July 2005
Exclusions	None reported
Population	n=944,078 newborns
Intervention/test	Newborn screening for MMA, PA and other conditions through first and second tier tests using tandem mass spectrometry (MS/MS) on dried blood spot. <u>Cutoff values:</u> Initially when the pilot started in 2002 the cut-offs were set at about mean + 4SD, with

	mean based on 2,000 samples. These were modified over time as more samples were analysed and more clinical data was available. For 2003-2004 the cutoff values were as below: 'Diagnostic' cutoff values: C3 >9.0 µM And C3/C2 >0.15 Borderline cutoff values: C3 >4.82 µM And C3/C2 >0.15 (Elevated C3/C2 ratios alone are not considered significant). If an elevated result was obtained, the test was repeated on the same sample. For confirmed borderline results a second sample was requested for testing, the exact timing of this sample was not reported. If this was also borderline the same procedure as for diagnostic cut-offs was followed. For confirmed results above the diagnostic cutoff the infant's primary care physician was contacted with results and recommendations so that the infant could be referred to a metabolic centre as needed, ideally within 24-48 hours.
Comparator	Confirmatory tests included a repeat of the MS/MS screen, UOA, and PACYLC, serum B12, and plasma homocysteine. Enzyme and mutation analyses were carried out where available and approved by reimbursers. Result of the latter tests was not needed to start treatment, as the aim was to start treatment within 10-14 days of birth.
Results/outcomes	<u>Screening test performance:</u> <u>For 2003-2004:</u> MMA True positives (TP): 5 (10 for entire period: 3 MUT0, 2 had "very mild variants with persistent elevation of methylmalonic acid in blood and urine", and 5 CblC; 1 infant with CblC died) MMA False positives (FP): 411 (if only one elevated borderline value was considered), 4 (if only repeat borderline values were considered), 4 (if only diagnostic cutoff considered) MMA True negatives (TN): Not reported or calculable MMA False negatives (FN): Not reported for 2003-2004 specifically. There was 1 false negative over the whole period (CblA, mild, presented at 5 months during intercurrent illness, reported to be developing normally at the time of the report; screening values had been below cut-offs and remained so on re-testing of the sample, and these were lowered as a result) MMA Positive predictive value (PPV): 1.20% (if only one elevated borderline value was considered), 55.56% (if only repeat borderline values were considered), 33.33% (if only

	diagnostic cutoff considered) MMA sensitivity: Not reported or calculable for 2003-2004 (overall 90.9% [RC]) MMA specificity: Not reported or calculable Incidence of MMA and PA: MMA: 1 in 90,000 PA: 1 in 300,000
Comments	Limited information on false positives (for 2003-2004 only) and no information on when the false negative was missed were provided. This limited the analyses of test performance which were possible. No systematic follow up of screen negatives or alerting system for false negatives was reported, so they may be under ascertained. However, some false negatives were identified.

Appendix number	11
Relevant criteria	KQ1, KQ3
Publication details	Lund AM, Hougaard DM, Simonsen H, et al. Biochemical screening of 504,049 newborns in Denmark, the Faroe Islands and Greenland--experience and development of a routine program for expanded newborn screening. Mol Genet Metab. 2012;107(3):281-93.[18]
Study details	Screening programme evaluation from Denmark, the Faroe Islands and Greenland
Study objectives	To present results of expanded newborn screening programme data from a 9 year period
Inclusions	Newborns born in Denmark, the Faroe Islands and Greenland between 1st February 2002 and 31st March 2011 whose parents accepted NBS
Exclusions	Non-participation in newborn screening (n=82,930)
Population	n=504,049 newborns
Intervention/test	Newborn screening for MMA, PA and other conditions using tandem mass spectrometry (MS/MS) on dried blood spot collected by heel prick at 4-9 days of age (trial period 2002-2009; median age 5 days) or 2-3 days of age (routine screening period 2009-2011; median age 2.5 days). Results were available 2-7 days (trial period) or 2-3 days (routine period) after samples received by the screening lab. Cutoff values: These were initially set based on literature, in-house experience on stores DBS samples from affected individuals and statistical assessment of reference cohorts. Cut-offs were adjusted over time to optimise screening performance and because of changing analytical method (different MS/MS machines and analysis kits used over time). The main change was after the trial period where samples were collected earlier and an underivatised screening method was used rather than derivatised. Performance of the

	screening lab was monitored by the CDC Newborn Screening Quality Assurance Program. Trial period: Primary: C3 >5.1 µM Secondary: C3/C2 >0.35 or C4DC >0.4U Routine period: Primary: C3 >6.0 µM Secondary: C3/C2 >0.25 Confirmatory tests are reported below as the 'Comparator' for the screening test results as they formed the gold standard against which false and true positives and negatives were judged.
Comparator	Confirmatory tests (diagnostic 'gold standard'): DBS with values above cut-offs were reanalysed in duplicate, and if abnormality was confirmed the diagnosis was confirmed or ruled out through testing of urine organic acids (UOA), plasma acylcarnitines (PACYLC), plasma amino acids (PAA), DNA, and propionate incorporation into fibroblasts. If the disease could not be confirmed in these tests the screening result was considered a false positive. False negatives were also recorded where a child had had screened negative, but was later identified as being affected by population screening of through affected families. Historical cohort: Infants who had clinical diagnoses of the disorders during the study period, but who were born before the expanded NBS panel was introduced, between January 1st 1992 and December 31st 2001, were used as a comparator in terms of presentation. The screening centre also carried out diagnosis and treatment of all children from the participating areas, so the majority or all of affected children diagnosed in this period were believed to have been identified.
Results/outcomes	**KQ3:** Screening test performance: True positives (TP): 3 MMA False positives (FP): 51 (1 premature; includes 2 PA) True negatives (TN): 503,992 False negatives (FN): 3 (1 with vitamin-B12 non-responsive MMA of unknown cause, 2 with MMA caused by methylmalonyl CoA epimerase deficiency) MMA/PA false positive rate: 0.0097% MMA/PA Positive predictive value (PPV): 9% MMA sensitivity: 50% [RC] MMA specificity: 99.99% [RC] MMA PPV: 5.56% [RC]

	KQ1: <u>Presentation and outcome of infants identified or missed by screening:</u> • 1/3 (33.3%) of MMA cases presented clinically before screening results were available (presented at 2 days of age with metabolic decompensation due to MMA mut$^-$) • 2/3 (66.7%) of those with MMA (cblA and cblB asymptomatic, symptomatic case was mut$^-$) • 100% (3/3) of infants with MMA did not have significant sequelae (were well with normal psychomotor development and developmental age) at last follow up - timing of last follow up was not reported and no formal cognitive testing was carried out • There were no deaths among the children identified with MMA through screening • One of the false negative cases (MMA of unknown cause) was identified after unexplained death of a sibling (whether he had symptoms himself was not reported), and the other two were siblings – the older of whom had metabolic decompensation at age 2 years. <u>Missed cases:</u> 1 case of transcobalamin II deficiency was identified among the infants whose parents refused screening. This disorder was not one being screened for but methylmalonic acidaemia was present as a feature of the condition. <u>Historical cohort:</u> 1 case of MMA was identified, who presented at age 2 days with metabolic decompensation. There were 674,754 births in the period. <u>Incidence of MMA/PA:</u> Screening cohort: 1 in 100,809 in those receiving screening, and 1 in 65,219 if the case in the screening refusal group included Historical cohort: 1 in 61,341
Comments	Although there was not systematic follow up of the screen negative infants, the fact that diagnostic tests and management of metabolic disorders for all three regions seemed to be carried out at the centre conducting the study, increases likelihood that missed cases would be detected and also known cases in the historical period assessed. In both periods individuals who died from or with the conditions without having ever been screened or diagnosed would have been missed. Exact treatment received by affected individuals was not reported. The authors noted that the high false positive rate could be reduced by a second tier MS/MS assessment of methylmalonic acid, 2-methylcitric acid, and total homocysteine in dried blood spots, but this would further delay availability of screening results. They suggest that before introducing such an assessment the basis for inclusion of MMA and PA

	in the screening panel should be reviewed. The incidence of the conditions did not differ between the screening and non-screening time periods, suggesting that additional cases might not be being detected by screening. However, given the rarity of these individual conditions, even larger numbers of births might be needed to be able to detect differences.

Appendix number	12
Relevant criteria	KQ1, KQ3
Publication details	Schulze A, Lindner M, Kohlmuller D, et al. Expanded newborn screening for inborn errors of metabolism by electrospray ionization-tandem mass spectrometry: results, outcome, and implications. Pediatrics. 2003;111(6 Pt 1):1399-406.[20]
Study details	Screening programme evaluation from Germany
Study objectives	To assess the impact of an expanded newborn screening programme on detection of IEM in one region in Germany over a 42 month period, and to assess outcome of those diagnosed with IEM
Inclusions	Newborns born in Baden-Wurtemberg, Germany between April 1998 and September 2001 whose parents accepted NBS
Exclusions	Non-participation in newborn screening (n=16,200)
Population	n=250,000 newborns
Intervention/test	Newborn screening for MMA, PA and other conditions using tandem mass spectrometry (MS/MS) on dried blood spot collected by heel prick at 3-7 days of age (median 5 days; 0.88% carried out before day 1, and 1.65% after day 7). Preterm infants had a second sample taken after 14 days of age. <u>Cutoff values:</u> Cutoff set at 99.5th percentile of values from MS/MS on DBS from 10,000 healthy neonates. These were reported as: C3 >6.8 µM (MMA and PA) C3/C0 >0.19 (MMA and PA) C3/C2 >0.39 (MMA and PA) Methylmalonylcarnitine >1 (MMA only) (analyte likely to be C4DC, containing methylmalonyl and succinylcarnitine) Repeat analysis was carried out on samples with values above the cutoff for one or more analytes (i.e. a re-test of the same blood spot). If the 1st test was >30% over the cutoff and the 2nd test was normal, a 3rd test was carried out and the mean of the three values used. Samples where both tests were positive for any analyte were taken to the next step:

	assessment by an experienced metabolic disease specialist. The specialist used a rating system based on level of deviation from normal and overall analyte profile to decide if the sample was a screen positive. The system was optimised based on measurements of infants with known disorders, reports in the literature, previous screening test performance (sensitivity and specificity), and assessor experience. A second DBS was requested from screen positives, or the infant was referred for hospital admission if avoidance of delay was felt to be essential for their wellbeing (conditions falling into this latter category not reported). Confirmatory tests are reported below as the 'Comparator' for the screening test results as they formed the gold standard against which false and true positives and negatives were judged.
Comparator	Confirmatory tests (diagnostic 'gold standard'): Diagnosis was confirmed or ruled out through testing of UOA (MMA: methylmalonic acid, PA: tiglylglycine, 3-hydroxypropionic acid, methylcitrate), plasma testing (MMA: ammonia, homocysteine; PA: ammonia), and enzyme activity in fibroblasts. Monthly questionnaires were sent to all pediatric hospitals and metabolic centres in Germany to identify any children with IEM missed by the screening programme (false negatives).
Results/outcomes	**KQ3:** <u>Screening test performance:</u> MMA True positives (TP): 4 MMA (1 mut, 2 cblC/D, 1 suspected mut with screening result confirmed on recall but diagnosis remaining uncertain because diagnosis difficult to achieve or person lost to follow up [not specified which in this case], those who were not lost to follow up did not show any symptoms (also 1 PA, counted as a false positive here) MMA False positives (FP): 206 (207 if unconfirmed case counted as a false positive)[RC] MMA True negatives (TN): 249,790 [RC] MMA False negatives (FN): 0 MMA Positive predictive value (PPV): 1.90% [RC] (1.43% if unconfirmed case counted as a false positive) MMA sensitivity: 100% MMA specificity: 99.92% **KQ1b:** <u>Presentation and outcome of infants identified by screening:</u> • 2/3 confirmed MMA cases (both cblC/D) were symptomatic before the time of the screening result (2/4 confirmed or suspected cases) • 1/3 confirmed MMA cases had been diagnosed before the screening result Details of symptoms were not provided.

	Incidence of MMA: 1 in 62,500 if unconfirmed case counted as a true positive [RC] 1 in 83,333 if unconfirmed case counted as a false positive [RC]
Comments	Although there was not systematic follow up of all of the screen negative infants, the fact that monthly questionnaires were sent to centres where the infants might be treated (paediatric hospitals and metabolic centres), increases likelihood that missed cases would be detected. This pilot explicitly used a rating system based on various criteria (including cut-offs) and assessed by a metabolic specialist in order to assign screen positives, rather than based on the cutoff value alone. To what extent results differed from what would have been obtained just using set cut-offs alone was not investigated. A later paper[30] reported that due to an unfavourable balance of sensitivity and specificity, disorders of propionate metabolism assessed by elevated C3 levels on NBS were excluded from the conditions recommended to be screened for nationally in German guidelines from 2005. The paper aimed to develop a statistical approach to new parameter combinations to give improved specificity and 100% sensitivity for disorders of propionate metabolism (not reported here as it was not an actual screening programme).

Appendix number	13
Relevant criteria	KQ3
Publication details	Lim JS, Tan ES, John CM, et al. Inborn Error of Metabolism (IEM) screening in Singapore by electrospray ionization-tandem mass spectrometry (ESI/MS/MS): An 8 year journey from pilot to current program. Mol Genet Metab. 2014;113(1-2):53-61.[29]
Study details	Screening programme evaluation from Singapore
Study objectives	To present results of a newborn screening programme from an 8 year period
Inclusions	Newborns born in Singapore between July 2006 and April 2014 whose parents accepted NBS
Exclusions	Non-participation in newborn screening
Population	n=177,267 newborns
Intervention/test	Newborn screening for MMA, PA and other conditions using tandem mass spectrometry (MS/MS) on dried blood spot collected by heel prick at over 1 day of age (mean 47.9 hours, about 2 days). Sample transport took a mean of 0.92 days, and results were available within a mean of 1.64 days (trial period) or 3.8 days (routine period, due to 2nd tier testing). Preterm infants had heel prick at the time of birth and repeated screens at 2 and 4 weeks of age. Cutoff values: In the pilot period (2006-2009) these were initially set at the 99th percentile of local newborn population values. From 2010 these were adjusted to set values

	between the 99th percentile of the newborn population and the 5th percentile of disorder population values and using the R4S data as a guide. Cut-offs in this period were: C3 ≥5.50 µM C3/C2 >0.20 C3/C16 >2.0 C3/Met >0.20 A high false positive rate as a result of elevated C3 levels prompted introduction of a 2nd tier liquid chromatography MS/MS test for methylmalonic acid. If C3 or C3/C2 ratios were elevated the level of methylmalonic acid was tested on DBS. If methylmalonic acid level was >3.0 µM the screen was considered positive for MMA. If this test was negative but C3>9 µM, the infant was investigated for PA. The authors reported that they were investigating another second tier test for total homocysteine, methylmalonic acid, and methylcitric acid to improve their MMA screening. They note that this test would cover some cobalamin disorders, homocystinuria and PA as well as MMA. Performance of the screening lab was monitored through the CDC Newborn Screening Quality Assurance Program and ERNDIM quality assurance programme, and the lab was a member of the R4S collaboration. DBS with 'significantly abnormal' values were reanalysed, and a repeat sample requested for borderline values. Confirmatory tests are reported below as the 'Comparator' for the screening test results as they formed the gold standard against which false and true positives and negatives were judged.
Comparator	Confirmatory tests (diagnostic 'gold standard'): diagnosis was confirmed or ruled out through testing of UOA (particularly looking for methylcitric acid), and PACYLC.
Results/outcomes	<u>Screening test performance:</u> Trial period: 61,313 newborns tested and 2 true positives – 1 CblC/D case, 1 MUT/CblA/B case (specific gene mutated not reported); no false negatives reported in the paper Routine period (2010 onwards, includes 1st and 2nd tier tests): True positives (TP): 3 cases MMA (0 cases PPA) False positives (FP): 667 for 1st tier test only; 16 for 1st & 2nd tier test True negatives (TN): 111,619 for 1st tier test only; 112,270 for 1st & 2nd tier test False negatives (FN): None reported

	Specificity [RC]: 99.4% with 1st tier test only; 99.98% with 1st & 2nd tier test Sensitivity: 100% PPV [RC]: 0.45% for 1st tier test only; 15.8% for 1st & 2nd tier test (NB The study counted 2 cases of maternal vitamin B12 deficiency identified through elevated C3 levels in newborn DBS as true positives, these have not been counted as true positives here). No PA cases were detected. <u>Presentation of infants not screened:</u> • 2 infants who had not been screened presented clinically with symptoms (age NR) and were diagnosed with MMA. "Most" of the 6 children presenting clinically with MMA or other inborn errors of metabolism were described as having irreversible developmental and neurological complications, but no further details were given. <u>Incidence:</u> CblC/D: 1 in 59,100 MUT/Cbl A, B: 1 in 88,600
Comments	There was no systematic follow up of the screen negative infants therefore false negatives may have been missed.

Appendix number	14
Relevant criteria	KQ3
Publication details	la Marca G, Malvagia S, Casetta B, et al. Progress in expanded newborn screening for metabolic conditions by LC-MS/MS in Tuscany: update on methods to reduce false tests. J Inherit Metab Dis. 2008;31 Suppl 2:S395-404.[28]
Study details	Screening programme evaluation from Italy
Study objectives	To present results of a newborn screening programme from a 6 year period
Inclusions	Newborns born in Tuscany between January 2002 and 2008 whose parents accepted NBS. The pilot period ran in 3 provinces from 2002-October 2004, and expanded to all Tuscany after this.
Exclusions	Non-participation in newborn screening
Population	n=160,000 newborns
Intervention/test	Newborn screening for MMA, PA and other conditions using liquid chromatography tandem mass spectrometry (MS/MS) on dried blood spot collected by heel prick at 2-3 days of age. Preterm infants had repeat samples taken at 15 and 30 days and transfused newborns had

	a repeat sample 7 days after end of transfusion. Cutoff values: In the pilot period (2002-2004) the cut-offs were initially set at the mean +2SD, and adjusted for specific metabolites (including C3) as needed to reduce false positive results. C3 had high recall and false positive rates, so the cutoff was raised to mean +4SD, ratios of other metabolites used as secondary markers, and second tier tests introduced for those falling between the +2SD and +4.5SD cut-offs. Cut-offs were: Primary marker: C3 >3.3 µM (initially), then raised to >5.65 µM, and then 2nd tier testing introduced for those between these levels Primary marker for MMA and PA, also a secondary marker for holocarboxylase synthetase deficiency (HCSD) – the newborn form of MCD Secondary markers: C4DC >0.54 µM (specific for MMA) Gly >721 µM (MMA and PA) C3/C0 >0.13 (MMA and PA) C3/C4 >12.5 (MMA and PA) C3/C16 >1.6 (MMA and PA) A high false positive rate as a result of elevated C3 levels prompted introduction of a 2nd tier test for methylmalonic acid and 3-hydroxypropionic acid on the bloodspot. This was used if C3 levels were above 5.65 µM or between 3.3 and 5.65 µM. The cut-offs used were not reported. Whether external quality assurance programs were used was not reported. For conditions where there was risk of acute metabolic compensation (which would be likely to include MMA and PA) babies with a positive screening result were recalled immediately by a metabolic specialist for clinical examination and confirmatory tests. In conditions without risk of acute metabolic compensation a second DBS was requested and MS/MS repeated to see if abnormal result was confirmed. Confirmatory tests are reported below as the 'Comparator' for the screening test results as they formed the gold standard against which false and true positives and negatives were judged.
Comparator	Confirmatory tests (diagnostic 'gold standard'): diagnosis was confirmed or ruled out through testing of UOA and PACYLC (exact metabolites not specified). Whether DNA testing was used was not reported.
Results/outcomes	Screening test performance: True positives (TP): 2 cases MMA (1 MUT, 1 CblC/D), 2 cases PA. The study also described 1 maternal case of MMA CblC identified as a result of elevated C3 in newborn DBS. The study did not explicitly state whether the infant was affected or not, but as it was described as a maternal case of CblC it seems unlikely that the child was affected.

	False positives (FP): Not reported. In 2004-2006 there were 124 C3 positive tests, but whether these included the true positives was not reported, so false positive rate could not be calculated. False negatives (FN): 2 (1 MMA CblC; 1 case of transcobalamin II defect, a disorder in which MMA is a feature, but was not being screened for). Both had C3 values above 3.3 µM, but were missed when the 5.65 µM cutoff for screen positives was being used Specificity: Not reported/calculable Sensitivity: Not reported/calculable PPV : Not reported/calculable <u>Incidence of MMA/PA:</u> Not reported
Comments	There was no systematic follow up of the screen negative infants therefore false negatives may have been missed. The lack of information about total positive tests over the 6 year period limits ability to calculate measures of test performance. The sensitivity figure which was calculated is for the entire period; methods changed during this period and results may not be representative of the individual methods used. Whether the study was a member of external quality assurance programme such as the CDC program was not reported, although CDC external standard blood spots were used every day to check machine performance.

Appendix number	15
Relevant criteria	KQ3
Publication details	Shigematsu Y, Hata I, Tajima G. Useful second-tier tests in expanded newborn screening of isovaleric acidemia and methylmalonic aciduria. J Inherit Metab Dis. 2010;33(Suppl 2):S283-8.[19]
Study details	Screening programme evaluation from Japan
Study objectives	To report on the results of a pilot newborn screening programme over about 3 years and 'selective screening' over 2 years
Inclusions	Newborns born in Japan between January 1997 and July 2001
Exclusions	None reported
Population	N=102,200 newborns, n=164 children undergoing testing due to clinical presentation ('selective screening')
Intervention/test	Newborn screening for MMA, PA, and other conditions using MS/MS. 'Some' DBS samples were collected at 5-6 days of age, the exact timing of the others was not reported (the 2 MMA cases had samples collected at 5 and 8 days). Some samples were reported to be being assessed retrospectively. Age at careening result was not reported.

	Patients presenting with a variety of symptoms which could be linked to metabolic disorders were also tested, this was referred to as 'selective screening' and is not reported further here. Their ages were unclear. <u>Cutoff values:</u> C3/C2 >0.25 (mean + 6 SD) C3 levels were reported for cases but it was not clear whether this was considered in screening, and no cutoff value was reported.
Comparator	Confirmatory tests were not explicitly reported. Urinary methylcitrate levels at 18 to 25 days of age, and residual enzyme activity were reported for cases with PA, suggesting that these were used as part of the diagnostic workup for that condition.
Results/outcomes	**KQ3:** True positives: 2 MMA (1 mut, 1 vitamin B12 responsive; not further specified) False positives: 60 (includes 5 with PA)[RC] (based on a recall rate of 0.06%) True negatives: 102,138 [RC] False negatives: None reported Sensitivity: 100% Specificity: 99.94% [RC] PPV: 3.23% Incidence: PA: 1 in 20,440 [RC] MMA: 1 in 51,100 [RC] **KQ1b:** One of the babies with MMA (mut deficiency) had symptoms at 8 days when the first DBS was collected, and the other (with a vitamin B12 responsive MMA) was asymptomatic 'before a high [methylmalonic acid] level in the first DBS collected at 5 days was observed. No details of the symptoms were provided.
Comments	Some aspects of the methods of the screening programme and study were not well reported (e.g. exact age at screening, what number of samples were being assessed retrospectively etc.) No systematic approach to false negative detection was reported, so some may be missed, and sensitivity is likely to be over-estimated.

Appendix number	16
Relevant criteria	KQ1, KQ3
Publication details	Cheng KH, Liu MY, Kao CH, et al. Newborn screening for methylmalonic aciduria by tandem mass spectrometry: 7 years' experience from two centers in Taiwan. J Chin Med

	Assoc. 2010;73(6):314-8.[22]
Study details	Screening programme evaluation, Taiwan
Study objectives	To assess whether an immediate accurate diagnosis of MMA can be made using initial laboratory data in newborns
Inclusions	Newborns screening positive for raised C3 levels in newborn dried blood spot screening at 2 centres in Taiwan from January 2002 to December 2008.
Exclusions	None reported
Population	n=598,522 newborns
Intervention/test	DBS samples were obtained at 48-72 hours of life and 24 -48 hours after feeding. Cut-offs for MMA were: C3 borderline: >7 µM (1 centre), ≥6 µM (1 centre) C3 positive: ≥10 µM (2 centres) C3/C2 borderline: ≥0.2 (1 centre), >0.25 (1 centre) C3/C2 positive: >0.5 (2 centres) A borderline result prompted a repeat sample and test, and if confirmed resulted in referral. A positive test resulted in immediate referral. Once referred, confirmatory tests included plasma ammonia, PAA, UOA, liver function and blood gas tests. If methylmalonic acid was identified in urine, then MCM enzyme activity was tested. True positives (n=7, all *MUT*^{0}) were those with MMA in urine and reduced MCM enzyme activity.
Comparator	In this study confirmed cases of MMA (true positives) identified through screening were compared with screening false positives. False positives (n=15) had a positive screening result, but normal urine organic acids, and no clinical signs of MMA.
Results/outcomes	There was complete overlap in the C3 values seen in true and false positives, while there was no overlap in the C3/C2 values, and the means in the 2 groups were significantly different. • C3: MMA range 7.7 to 19.18 µM, mean 13.34 µM vs. FP: 8.2 to 22.6 µM, mean 11.55; p=0.217 • C3/C2: MMA range 0.55 to 1.18, mean 0.77 vs. FP: 0.12 to 0.387, mean 0.26; p<0.0001 True positives and false positives also differed significantly in the mean values of the markers used as part of diagnostic work up: higher plasma ammonia (p<0.0001), lower blood gas pH (p=0.029), lower bicarbonate (p=0.019), higher aspartate aminotransferase (AST, p=0.005), higher alanine aminotransferase (ALT, p=0.081). There was overlap in the values seen for these markers between MMA true positives, false positives and normal values, other than ammonia (MMA range: 239-695 µg/dL; normal range: 90-150 µg/dL; false positive range: 35-125 µg/dL). They calculated that if they had used the existing C3/C2 ratio cut-offs as the primary parameter for referral this would have given 100% sensitivity and a PPV of 70%. If they

	used what was reported as the German C3/C2 ratio cutoff of >0.39, this would have given a sensitivity of 100% and PPV of 100%. On this basis they suggested re-evaluating whether C3/C2 ratio should replace C3 as the primary marker. They concluded that a C3/C2 ratio of >0.4 or ammonia levels > 200 µg/dL should be highly suspected of having MMA. **KQ1b:** Symptoms in MMA screen detected cases: All 7 true positives had metabolic decompensation when they were admitted for confirmatory testing. Their clinical manifestations were reported to include poor activity, poor appetite, vomiting, and shortness of breath.
Comments	It was not clear whether the false positives could have included cases of PA or other conditions (e.g. MCD), or carriers of MMA mutations. The false positives had no clinical symptoms. The cutoff values reported for the 2 centres differ from those reported in Niu et al. 2010[21] for an overlapping period in these centres. This may be due to changes in cut-offs over time. The population in this study may overlap with that in Niu et al. 2010[21] It was unclear whether the true positives with metabolic decompensation all had symptoms or whether some only showed biochemical signs.

Appendix number	18
Relevant criteria	KQ4
Publication details	Barends M, Pitt J, Morrissy S, et al. Biochemical and molecular characteristics of patients with organic acidaemias and urea cycle disorders identified through newborn screening. Mol Genet Metab. 2014;113(1-2):46-52.[34]
Study details	Case series with historical control group, Australia
Study objectives	To identify markers that may assist in predicting the need for treatment in infants with classical organic acidaemias (OAs) and urea cycle disorders (UCDs) identified through newborn screening.
Inclusions	All patients with classical OA (MMA, PA, isovaleric acidaemia [IVA]) identified through the Victoria screening programme, from February 2002 to January 2014, or diagnosed clinically between 1990 and January 2002. (Methods and results for UCDs are not considered here)
Exclusions	None stated.
Population	n= 1 MMA (MUT) n=2 MMA (1 MUT, other not specified)

Intervention/test	Newborn MS/MS screening.
Comparator	Clinical detection of cases.
Results/outcomes	The 1 MMA case detected by screening was treated with a low protein diet, carnitine (100 mg/kg/day), and a sick day regime (not further described). The 2 clinically detected cases had these treatments plus metronidazole citravescent. Further details of treatment regimens were not given. The infant detected by screening was reported to present clinically in the first few days of life (not reported whether this was before screening results available). He had: • mild intention tremor for a short period but reported as asymptomatic after the newborn period • 11 admissions, only one due to metabolic decompensation, which was within the first 5 days of life, this was the only admission in the first year of life, and there were 2 in the second year of life • age appropriate development • some difficulties with higher order neuropsychiatric skills at age 8 years and 7 months (processing speed, fine motor skills, divided attention) • normal renal function at age 10 years and 9 months • follow up to age 11 years 6 months The 2 clinically diagnosed MMA patients both had over 10 admissions due to metabolic decompensation in both the first and second years of life.
Comments	The causative mutations of only 2 of the cases were identified (both had the same MUT mutations), and it was unclear whether these conveyed complete or partial loss of Mut function. Results may be confounded by differences in treatment between the two time periods. While the components of the MMA treatment regimens were reported to be the same, details of how these and other treatments were administered may have varied. In addition, it was unclear to what extent there was adherence to the treatment regimen in the 3 cases. Alert guidelines and decisions relating to prophylactic admissions were reported to be the similar in the two periods. The infant detected by screening presented in the first few days of life, and was hospitalised before 5 days of age with metabolic decompensation. Given that collection of bloodspot for screening took place at 2-3 days of life, presentation was likely to have been before results were available, although this was not explicitly stated. Very little information was provided relating to the clinically detected cases (including duration of follow up, age at presentation, or results of any developmental or neuropsychiatric testing). The small number of cases and the limitations above make it difficult to draw firm conclusions about differences in the outcome of screen detected and clinically detected

	cases.

Appendix number	19
Relevant criteria	KQ4
Publication details	Wilcken B, Haas M, Joy P, et al. Expanded newborn screening: outcome in screened and unscreened patients at age 6 years. Pediatrics. 2009;124(2):e241-8.[32]
Study details	Case series with historical control group, Australia
Study objectives	To study the clinical outcome at age 6 years of children identified as having an inborn error of metabolism (IEM) through newborn screening or clinical diagnosis.
Inclusions	Children with amino acid, organic acid, or fatty acid metabolism detectable through newborn MS/MS screening, and identified through screening (1998 to 2002) or clinical presentation (1994 to 2002, 3 from the period where screening was available but infants unscreened).
Exclusions	None reported
Population	n=3 MMA (2 MUT or CblA, B; 1 Cbl C) detected in screened population and 16 MMA (10 MUT or CblA, B; 6 CblC) cases detected clinically
Intervention/test	Newborn MS/MS screening.
Comparator	Clinical detection.
Results/outcomes	One MMA case missed by screening (false negative, Cbl C) died by age 6 years (1/3, 33%). One MMA case identified through clinical presentation also died by age 6 years (1/16, 6.25%). One additional MMA case identified through clinical presentation died between age 6 and 10 years (1/16, 6.25%).
Comments	This study also provided information on intellectual handicap, school placement, and physical score for screened and unscreened groups with IEM but did not provide separate outcomes for MMA cases. The period in which screening was offered was shorter than the period when screening was not offered, therefore follow-up was longer and number of cases higher for the latter. Results may be confounded by differences in treatment between the two time periods. Less than half of the infants in the period 1998-2002 were screened, so there were still some clinically detected cases in this period, but this represented only 3 out of the 16 cases detected clinically. Details of treatments were not provided, other than that a sick day regimen was provided to parents, as well as instructions to telephone metabolic physician on call for advice during illness and a letter to expedite triage and management if the child presented at accident and emergency. Other than death, no other outcomes were reported specifically for children with MMA. The child in the screened group who died had been missed by screening.

	The small number of cases and the limitations above make it difficult to draw firm conclusions about differences in the outcome of screen detected and clinically detected cases.

Appendix number	20
Relevant criteria	KQ4
Publication details	Dionisi-Vici C, Deodato F, Roschinger W, et al. 'Classical' organic acidurias, propionic aciduria, methylmalonic aciduria and isovaleric aciduria: long-term outcome and effects of expanded newborn screening using tandem mass spectrometry. J Inherit Metab Dis. 2006;29(2-3):383-9.[31]
Study details	Comparative case series
Study objectives	To compare the natural history of patients with classical organic acidurias (including MMA) diagnosed on a clinical basis with those diagnosed through newborn MS/MS screening.
Inclusions	Clinically detected cases of MMA from a children's hospital in Italy from 1983, and cases detected by newborn screening in Australia and Germany (time period not stated).
Exclusions	Enter applicable exclusion criteria (study and participant criteria for SRs)
Population	N=15 clinically diagnosed MMA cases and 4 cases of MMA detected by screening. (The latter may include infants diagnosed due to affected older sibling).
Intervention/test	Newborn MS/MS screening.
Comparator	Clinical detection of cases.
Results/outcomes	<u>Clinically detected cases:</u> Among the 5 MMA cases with neonatal onset for which this information was presented: • 3/5 died (60%, 2 at about age 6 and one at about age 2 years), the 2 survivors were aged around 6 and 12 years • 4/5 (80%) had at least one coma, 5/5 had at least one metabolic decompensation, and 4/5 (80%) had multiple metabolic decompensations and the one who did not died at age 2 • 4/5 (80%) developed mental retardation at some stage (IQ≤79, or DQ≤74), the one who did not was currently aged around 12 years and had experienced multiple metabolic decompensations and one coma No information on the metabolic outcomes of the remaining 10 MMA cases was reported, including any late onset MMA cases. One MMA case (described as Mut^t) was reported as having 2 successful pregnancies with uneventful postpartum period. 2 MMA cases (2/15, 13%) were described as having basal ganglia stroke All MMA cases reaching the age of 6 years showed signs of chronic renal failure after this

	point. Screen detected cases: The 4 cases appeared to be aged around 2-3 years of age (3 cases) and around 6-7 years (1 case), outcomes included: • All 4 screen detected cases were symptomatic before the result of screening was known • None had died by last follow up • 1/4 (25%) had mild metabolic instability, and 1/4 (25%) had severe metabolic instability (neither term defined, not how this related to metabolic decompensation) • 2/4 (50%) had mild neurocognitive impairment (not defined) • 1/4 (25%) showed mild renal dysfunction (aged 40 months at last assessment) Incidence: The incidence of MMA and PA was not reported to be significantly different between the two groups (0.99 ± 0.35 per 100,000 in the clinically detected group and 1.04 ± 0.36 per 100,000 in the screen detected group) Results were not provided for MMA and PA separately.
Comments	It was unclear whether the cases described were all of the cases detected, or a subset. Whether both groups had similar assessment and monitoring of outcomes was unclear, as was the duration of follow up, which is likely to have been shorter for the screen detected cases. The time period in which the cases were detected was unclear, as the clinically detected group were diagnosed from 1983, when MS/MS would not have been available until later, it is likely the screened and clinically diagnosed groups are not contemporaneous. It was also unclear whether the treatment regimens were the same across the countries. These limitations and the small number of cases make it difficult to draw firm conclusions about differences in the outcome of screen detected and clinically detected cases, particularly relating to the impact of treatment.

Appendix number	21
Relevant criteria	KQ4
Publication details	Waisbren SE, Albers S, Amato S, et al. Effect of expanded newborn screening for biochemical genetic disorders on child outcomes and parental stress. Jama. 2003;290(19):2564-72.[33]
Study details	Prospective comparative case series (comparing inception cohorts), USA
Study objectives	To compare newborn identification by expanded screening with clinical identification of

	biochemical genetic disorders.
Inclusions	Children identified with biochemical genetic disorders (including MMA and PA) through expanded newborn screening in 3 states (2 in New England plus Pennsylvania), and children identified clinically in any of the 6 New England states between January 1999 and June 2002, and assessed by December 2002. Parents of diagnosed children were invited to participate between 5 and 30 months after diagnosis. 82% of the newborn screening group (those not enrolled reported to have the same diagnoses as those enrolled, except for 2 disorders [not MMA or PA]) and 97% of the clinically identified group agreed to participate. Children were given standard medical examination as well as having their medical records obtained for assessment.
Exclusions	Infants who died before enrolment (5 from screened group – none with MMA or PA), children from centres which failed to obtain internal approval for the study (43 families of clinically identified children, diagnoses NR) and those who could not be contacted (10 children in the newborn screening group, diagnoses NR).
Population	N=2 MMA cases in screened group and 2 MMA cases (0 CblC, others not defined) in clinically diagnosed group (1 CblC, other not defined) (Study also included 3 screened and 6 clinically identified cases of PA not described here).
Intervention/test	Newborn MS/MS screening.
Comparator	Clinical detection of cases.
Results/outcomes	• None of the children with MMA (screened or clinically detected) died during the study or needed neonatal intensive care before diagnosis • 1/2 (50%) of the children with screen-detected MMA and 1/2 (50%) clinically detected MMA cases (the Cbl C case) performed in the range of mental retardation on the Bayley Scales of Infant Development up to the age of 3 (not further defined) . No additional children aged over 3 years with MMA were reported to perform in the range of mental retardation, although it was unclear whether any of the children with MMA were this old.
Comments	The study did report other outcomes , but for children with metabolic disorders as a whole and not for MMA separately. Age at diagnosis and evaluation was lower for the newborn screened group than the clinically detected group as a whole (diagnosis: median 5 days [range 1 to 180] screened vs. 4 months [range 0.1 months to 5.9 years] clinically detected; evaluation: median 9 months [range 5 to 91] screened vs. 34 months [range 4 to 101] clinically detected, $p<0.001$ for both). These figures were not reported for MMA or PA specifically, therefore it is difficult to determine whether the children being compared with these diagnoses are comparable in terms of age. Follow up was reported to be short (exact duration not specified). Details of treatments received were not provided.

Appendix number	22
Relevant criteria	KQ5
Publication details	American College of Medical Genetics. Newborn Screening: Toward a Uniform Screening Panel and System. 2006[8]
Study details	Practice guideline, USA, developed by a multidisciplinary expert group (n=292) working with a steering committee and expert working groups. The group provided their opinions on the extent to which individual met criteria relating to: 1. The availability and characteristics of the screening test 2. The availability and complexity of diagnostic services 3. The availability and efficacy of treatments related to the conditions. The conditions were then ranked based on results and then the evidence base was assessed in depth using systematic reviews, internet searches, professional guidelines, clinical evidence, and cost/economic modelling and modelling. At least 2 experts assessed the data and level and quality of the evidence for each condition. The results of the evidence review and expert survey were considered by the ACMG along with over-riding principles, and other existing technology and condition specific recommendations to place the conditions into (a) core screening panel, (b) secondary targets (part of the differential diagnosis of a core panel condition, (c) not appropriate for newborn screening. The ACMG also assessed to what extent screening programmes met goals such as the availability of educational programmes, uptake of screening and follow up based on data from the National Newborn Screening and Genetics Resource Centre 2002 and experts. A brief cost-effectiveness analysis was also carried out.
Study objectives	1. To analyse the scientific literature on the effectiveness of newborn screening. 2. To gather expert opinion to outline the best evidence for screening for specified conditions and develop recommendations focused on newborn screening, including the development of a uniform condition panel. 3. To consider other parts of the newborn screening system that are critical to achieving the expected outcomes in those screened.
Inclusions	NA
Exclusions	NA
Population	Newborns
Intervention/test	Newborn screening for conditions including MMA
Comparator	NA
Results/outcomes	Final outcome was as follows: • Core panel: 29 conditions, including MMA caused by MUT mutations, as well as Cbl A,B (and also PA and MCD) • Secondary panel: 25 conditions, including MMA caused by Cbl C,D

	• Not appropriate for newborn screening: 27 conditions They decided to subdivide MUT, Cbl A,B, C, and D and PA since they had quite different natural histories and treatment options. The MUT form of MMA was the higher scoring primary target (i.e. identified by experts as a better candidate for screening), while CblA,B and PROP were lower scoring primary targets. The guideline noted that the forms of MMA caused by defects in adenosyl-cobalamin synthesis may overlap with dietary deficiencies. For MMA there was "credible evidence of less than ideal sensitivity with the current testing technology (affected cases with normal concentration when tested at birth) and specificity (relatively high rate of false-positive results, including cases with relatively high levels that are followed up by perfectly normal plasma acylcarnitine and urine organic acid profiles). They suggested that a second tier test to detect methymalonic acid in bloodspots could improve test performance. On the basis of the guideline, fact sheets confirmatory algorithms were developed which are keep updated on an ongoing basis.
Comments	This guidance was developed in the US, and relied on expert views as well as research evidence. The UK viewpoint may differ. The experts assessing the evidence on MMA differed in their rating of the level of evidence on the condition, with one rating it all as level 4 (derived from expert opinion, case reports, and reasoning from first principles; except for treatment of Mut deficiency, for which no rating was reported) while the other rated the evidence on testing and diagnosis as Level 1 for the test and diagnosis (derived from well-designed RCTs or diagnostic studies on relevant populations) and Level 2 for the condition and treatment (derived from RCTs or diagnostic studies with minor limitations; overwhelming, consistent evidence from observational studies). The ACMG note that some conditions have limited evidence, often because they are very rare. They also noted that for conditions with multiple forms with differing severity and age of onset, the decisions were based on the more severe and treatable forms of the conditions. Whether MMA was one of the conditions to which this statement applied specifically was not explicitly stated. The guideline had noted such differences between the different forms of MMA as a reason for subdivision of these conditions on the newborn screening panel list. It noted that there was there potential for bias in the assessment of some conditions, as without screening, the more severe forms are noticed first, biasing what is known about the effects of the condition. Until a large general population has been studied understanding of the performance of the screening test (in terms of the range of manifestations it identifies) is limited.

Appendix number	23
Relevant criteria	KQ1, KQ5

Publication details	Baumgartner MR, Horster F, Dionisi-Vici C, et al. Proposed guidelines for the diagnosis and management of methylmalonic and propionic acidemia. Orphanet J Rare Dis. 2014;9:130.[6]
Study details	European guideline, USA, developed using Scottish Intercollegiate Guideline Network (SIGN) methodology. A multidisciplinary guideline development group (GDG) (including representation from the UK) performed the systematic review (search date December 2011), drafted the guideline, discussed it with other GDG members and revised it based on this and other discussions with external stakeholders. Statements (recommendations) were graded: • A: based on the Level 1 evidence (meta-analyses, systematic reviews of RCTs, RCTs) • B: based on Level 2 evidence (case-control or cohort studies or systematic reviews of these) • C: based on level 3 evidence (non-analytic studies such as case reports and case series) • D: based on level 4 evidence (expert opinion)
Study objectives	To standardise the diagnosis, therapy and long-term management of MMA and PA in Europe based on the highest level of evidence available.
Inclusions	NA
Exclusions	NA
Population	Newborns
Intervention/test	Diagnosis and management of MMA and PA
Comparator	NA
Results/outcomes	70 recommendations are given for the diagnosis and management of MMA and PA. Those most relevant to the key questions of the current review are reported here. **KQ1a:** • Statement 8 (Grade C-D) Defects in different genes can cause isolated methylmalonic aciduria. The clinical phenotype is influenced by the underlying enzymatic defect (mut0, mut−, cblB, cblA and cblD-variant 2) and genotype (mut, MMAA, MMAB, MMADHC). • Statement 9 (Grade B) No clear-cut genotype-phenotype correlations have been found in PA. **KQ5:** • Statement 11 (Grade C-D) Newborn screening for MMA and PA is technically feasible. So far available data about outcome has not answered the question as to whether newborn screening in MMA/PA is of long term clinical benefit. • Statement 6 (Grade B-C) Determination of organic acids in urine and the acylcarnitine profile in blood are the most commonly used investigations to

	detect MMA and PA. Determination of amino acid concentrations may help in diagnosis and treatment. In addition total plasma homocysteine allows differentiation between the various types of MMA. • Statement 7 (Grade B/D) Enzymatic studies and/or molecular genetic analyses should be performed to confirm diagnosis (B). This is ideally done in specialized laboratories (D). • Statement 30 (Grade C-D) Response to vitamin B12 should be assessed in every MMA patient. For responders hydroxocobalamin should be used as long-term treatment. Doses of hydroxocobalamin have to be tailored individually depending on the clinical and biochemical results. Additional recommendations covered long term management, which included: • Dietary management, mainly low protein diet (exact levels reported to be guided by age, growth, metabolic stability and severity of condition) with PA precursor free amino acid supplements as needed to make up for any deficit. • L-carnitine supplementation to enhance propionyl group elimination and transform certain toxic metabolites into less toxic forms • Oral antibiotics continuously or intermittently to control intestinal propionic acid producing bacteria • Avoidance of certain drugs such as steroids and chemotherapy drugs **KQ6:** Statement 10 (Grade D) Prenatal testing in both diseases is feasible. Prior to testing, it is desirable that the index case has been confirmed biochemically and/or genetically, and the carrier status of the parents has been confirmed by mutation analysis.
Comments	The authors noted that the consensus recommendations, although based on best available evidence, often only represented expert opinion, and were meant to be followed flexibly, applying the practitioner's own experience, and taking into account the individual patient.

References

1. US National Library of Medicine. Methylmalonic acidemia [Internet].[updated July 2011; cited 21 April 15]. Available from: http://ghr.nlm.nih.gov/condition/methylmalonic-acidemia.
2. Orphanet. The portal for rare diseases and orphan drugs [Internet].[updated 20/05/2015; cited 21 May 2015]. Available from: http://www.orpha.net/consor/cgi-bin/index.php.
3. American College of Medical Genetics. Newborn screening confirmatory algorithm: C3 elevated (isolated) [Internet].[updated 2009; cited 21 May 2015]. Available from: https://www.acmg.net/StaticContent/ACT/Algorithms/Visio-C3.pdf.
4. Fernandez-Lainez C, Aguilar-Lemus JJ, Vela-Amieva M, et al. Tandem mass spectrometry newborn screening for inborn errors of intermediary metabolism: abnormal profile interpretation. Curr Med Chem. 2012;19(26):4511-22.
5. McHugh D, Cameron CA, Abdenur JE, et al. Clinical validation of cutoff target ranges in newborn screening of metabolic disorders by tandem mass spectrometry: a worldwide collaborative project. Genet Med. 2011;13(3):230-54.
6. Baumgartner MR, Horster F, Dionisi-Vici C, et al. Proposed guidelines for the diagnosis and management of methylmalonic and propionic acidemia. Orphanet J Rare Dis. 2014;9:130.
7. Dietzen DJ, Rinaldo P, Whitley RJ, et al. National academy of clinical biochemistry laboratory medicine practice guidelines: follow-up testing for metabolic disease identified by expanded newborn screening using tandem mass spectrometry; executive summary. Clin Chem. 2009;55(9):1615-26.
8. American College of Medical Genetics. Newborn screening: toward a uniform screening panel and system. 2006. Available from: https://www.acmg.net/ACMG/Publications/Practice_Guidelines_docs/NBS_report.aspx.
9. Horster F, Garbade SF, Zwickler T, et al. Prediction of outcome in isolated methylmalonic acidurias: combined use of clinical and biochemical parameters.[Erratum appears in J Inherit Metab Dis. 2009 Dec;32(6):762-3]. J Inherit Metab Dis. 2009;32(5):630-9.
10. Horster F, Baumgartner MR, Viardot C, et al. Long-term outcome in methylmalonic acidurias is influenced by the underlying defect (mut0, mut-, cblA, cblB). Pediatr Res. 2007;62(2):225-30.
11. Lerner-Ellis JP, Anastasio N, Liu J, et al. Spectrum of mutations in MMACHC, allelic expression, and evidence for genotype-phenotype correlations. Hum Mutat. 2009;30(7):1072-81.
12. Nogueira C, Aiello C, Cerone R, et al. Spectrum of MMACHC mutations in Italian and Portuguese patients with combined methylmalonic aciduria and homocystinuria, cblC type. Mol Genet Metab. 2008;93(4):475-80.
13. Morel CF, Lerner-Ellis JP, Rosenblatt DS. Combined methylmalonic aciduria and homocystinuria (cblC): phenotype-genotype correlations and ethnic-specific observations. Mol Genet Metab. 2006;88(4):315-21.
14. Merinero B, Perez B, Perez-Cerda C, et al. Methylmalonic acidaemia: examination of genotype and biochemical data in 32 patients belonging to mut, cblA or cblB complementation group. J Inherit Metab Dis. 2008;31(1):55-66.
15. UK Newborn Screening Programme Centre. Guidelines for Newborn Blood Spot Sampling [Internet].[updated February 2015; cited 21 May 2015]. Available from: http://newbornbloodspot.screening.nhs.uk/bloodspotsampling.

16. NHS Newborn Bloodspot Screening Programme. Standards for newborn blood spot screening [Internet].[updated August 2013]. Available from: http://newbornbloodspot.screening.nhs.uk/standards.
17. Karam PE, Habbal MZ, Mikati MA, et al. Diagnostic challenges of aminoacidopathies and organic acidemias in a developing country: a twelve-year experience. Clin Biochem. 2013;46(18):1787-92.
18. Lund AM, Hougaard DM, Simonsen H, et al. Biochemical screening of 504,049 newborns in Denmark, the Faroe Islands and Greenland--experience and development of a routine program for expanded newborn screening. Mol Genet Metab. 2012;107(3):281-93.
19. Shigematsu Y, Hata I, Tajima G. Useful second-tier tests in expanded newborn screening of isovaleric acidemia and methylmalonic aciduria. J Inherit Metab Dis. 2010;33(Suppl 2):S283-8.
20. Schulze A, Lindner M, Kohlmuller D, et al. Expanded newborn screening for inborn errors of metabolism by electrospray ionization-tandem mass spectrometry: results, outcome, and implications. Pediatrics. 2003;111(6 Pt 1):1399-406.
21. Niu DM, Chien YH, Chiang CC, et al. Nationwide survey of extended newborn screening by tandem mass spectrometry in Taiwan. J Inherit Metab Dis. 2010;33(Suppl 2):S295-305.
22. Cheng KH, Liu MY, Kao CH, et al. Newborn screening for methylmalonic aciduria by tandem mass spectrometry: 7 years' experience from two centers in Taiwan. J Chin Med Assoc. 2010;73(6):314-8.
23. Kronn D, Mofidi S, Braverman N, et al. Diagnostic guidelines for newborns who screen positive in newborn screening. Genet Med. 2010;12(12 Suppl):S251-5.
24. American College of Medicial Genetics. Newborn Screening ACT Sheet [Elevated C3 Acylcarnitine] Propionic Acidemia and Methylmalonic Acidemia [Internet].[updated 2012; cited 21 May 2015]. Available from: https://www.acmg.net/StaticContent/ACT/C3.pdf.
25. Kronn D, Mofidi S, Braverman N, et al. Diagnostic Guidelines for Confirmation of Screen-Positive Newborn Screening Results [Internet].[cited 06/05/2015]. Available from: http://www.wadsworth.org/newborn/nymac/docs/DX_Guidelines_2014-10-01.pdf.
26. Weisfeld-Adams JD, Morrissey MA, Kirmse BM, et al. Newborn screening and early biochemical follow-up in combined methylmalonic aciduria and homocystinuria, cblC type, and utility of methionine as a secondary screening analyte. Mol Genet Metab. 2010;99(2):116-23.
27. Frazier DM, Millington DS, McCandless SE, et al. The tandem mass spectrometry newborn screening experience in North Carolina: 1997-2005. J Inherit Metab Dis. 2006;29(1):76-85.
28. la Marca G, Malvagia S, Casetta B, et al. Progress in expanded newborn screening for metabolic conditions by LC-MS/MS in Tuscany: update on methods to reduce false tests. J Inherit Metab Dis. 2008;31 Suppl 2:S395-404.
29. Lim JS, Tan ES, John CM, et al. Inborn Error of Metabolism (IEM) screening in Singapore by electrospray ionization-tandem mass spectrometry (ESI/MS/MS): An 8 year journey from pilot to current program. Mol Genet Metab. 2014;113(1-2):53-61.
30. Lindner M, Ho S, Kolker S, et al. Newborn screening for methylmalonic acidurias--optimization by statistical parameter combination. J Inherit Metab Dis. 2008;31(3):379-85.
31. Dionisi-Vici C, Deodato F, Roschinger W, et al. 'Classical' organic acidurias, propionic aciduria, methylmalonic aciduria and isovaleric aciduria: long-term outcome and effects

of expanded newborn screening using tandem mass spectrometry. J Inherit Metab Dis. 2006;29(2-3):383-9.
32. Wilcken B, Haas M, Joy P, et al. Expanded newborn screening: outcome in screened and unscreened patients at age 6 years. Pediatrics. 2009;124(2):e241-8.
33. Waisbren SE, Albers S, Amato S, et al. Effect of expanded newborn screening for biochemical genetic disorders on child outcomes and parental stress. Jama. 2003;290(19):2564-72.
34. Barends M, Pitt J, Morrissy S, et al. Biochemical and molecular characteristics of patients with organic acidaemias and urea cycle disorders identified through newborn screening. Mol Genet Metab. 2014;113(1-2):46-52.