Newborn Blood Spot Screening for Tyrosinaemia Type 1 in the UK

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

Review group:	Julia Geppert
	Sian Taylor-Phillips
	Chris Stinton
	Karoline Freeman
	Hannah Fraser
	Sam Johnson
	Paul Sutcliffe
	Aileen Clarke
Correspondence to:	Dr Sian Taylor-Phillips
	Populations, Evidence and Technologies
	Division of Health Sciences
	Warwick Medical School
	University of Warwick
	Coventry CV4 7AL
Tel:	02476 575882
Email:	S.Taylor-Phillips@warwick.ac.uk

Date completed: 15/4/2016

Funding Acknowledgement

This research was commissioned by the UK National Screening Committee. Sian Taylor-Phillips, Chris Stinton, Hannah Fraser and Aileen Clarke are supported by the National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care West Midlands at University Hospitals Birmingham NHS Foundation Trust. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research, the UK National Screening Committee, Public Health England or the Department of Health. Any errors are the responsibility of the authors. The authors have no conflicts of interest.

Contents

List of Figures

List of Tables

List of Appendices

List of Abbreviations

AA	Amino acids
AC	Acylcarnitines
CDC	Centers for Disease Control and Prevention
CF	Cystic fibrosis
CHT	Congenital hypothyroidism
DBS	Dried blood spots
EPHPP	Effective Public Health Practice Project
FAH	Fumarylacetoacetate hydrolase, fumarylacetoacetase
GA1	Glutaric aciduria type 1
HCC	Hepatocellular carcinoma
HCU	Homocystinuria
IEM	Inborn errors of metabolism
IVA	Isovaleric acidaemia
LC	Liquid chromatography
MCADD	Medium-chain acyl-CoA dehydrogenase deficiency
MS/MS	Tandem mass spectrometry
MSUD	Maple syrup urine disease
NBS	Newborn blood spot
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NSQAP	Newborn Screening Quality Assurance Program
NTBC	2-(2-nitro-4-trifluoromethylbenzoyl)-1,3-cyclohexanedione (also known as Nitisinone or Orfadin®)
OLT	Orthotopic liver transplantation
OMIM	Online Mendelian Inheritance in Man
PHE	Public Health England
PKU	Phenylketonuria
PPI	Patient and Public Involvement
PPV	Positive predictive value
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PT	Proficiency testing
REA	*Rapid Evidence Assessment*
RHA	Regional Health Authority
SCD	Sickle cell disease
SUAC	Succinylacetone
TMS	Tandem mass spectrometry
TYR1	Tyrosinaemia type 1
UK NSC	UK National Screening Committee

Plain English Summary

The condition:

Tyrosinaemia type 1 (TYR1) is a rare, inherited condition which prevents the complete breakdown of tyrosine, an amino acid. Build-up of breakdown products can cause damage particularly to the liver, kidneys, and the nervous system. TYR1 can occur in two forms: early onset within the first months of life and chronic which is slower to develop. If it is untreated, TYR1 causes very severe illness and death from liver failure, or from other complications within the first 10 years of life.

The treatment:

The treatment consists of a drug called Nitisinone combined with a strict diet. Liver transplant is performed if patients do not respond to Nitisinone or develop liver cancer.

Screening and previous/current NSC UK recommendations:

Newborn screening has been suggested to identify children with TYR1 before they become ill in order to give early treatment to improve the health of the baby. The screening test would involve taking blood from the baby's heel, as part of the routine Newborn Blood Spot (NBS) Screening Programme. The blood is analysed for signs of TYR1 by measuring the levels of the toxic substance succinylacetone (SUAC).

The most recent review in 2014 concluded that newborn screening for TYR1 could be considered but more information would be required about the frequency of this condition in the UK, the feasibility of screening for TYR1, and the advantages of early treatment following screening versus later treatment following the onset of the illness. This review focuses on those areas. We searched for evidence published until September 2015.

Findings:

It is not known how many newborn babies are affected with TYR1 in the UK. The only data available for the UK are from two studies published in 1996 and 1998.

The evidence on the screening test shows that most of those with positive test results would have TYR1, but there is not enough information available following up babies who have had a negative screening test.

There is evidence that Nitisinone is an effective treatment. There is some evidence that children with TYR1 treated early might have better outcomes than those treated later, but this is from studies which are likely to be biased, and are not applicable to the current UK question. It is not clear whether the babies who have better outcomes from earlier treatment simply have a milder form of disease, or whether there are sub-clinical forms of disease which may be detected at screening.

Recommendation:

We need more research to help us find out

- How many babies are born with TYR1 in the UK each year and how many of them are already detected by current routes (blood spot screening for phenylketonuria [PKU], or testing of babies who have affected siblings)?
- How many babies who have a negative heel prick test according to low SUAC levels go on to develop TYR1?
- Whether babies with TYR1 detected by tyrosinaemia screening have a different form of disease than babies whose disease was identified once they developed symptoms or after following up a positive PKU screening test.
- Whether screening for TYR1 would result in a greater number of identified babies with TYR1 and whether all of these cases would benefit from treatment.

Executive summary

Introduction

Tyrosinaemia type 1 (TYR1) is a rare, inherited disorder of amino acid metabolism in which the complete breakdown of tyrosine is prevented, causing a build-up of tyrosine and its metabolites in the liver, kidneys and central nervous system. Babies with TYR1 have inherited an abnormal gene from both parents, which is much more likely when parents are from the same family (consanguinity). TYR1 can clinically present in two forms: an acute form which is characterised by early onset usually within the first months of life, and a chronic form which is slower to develop. If untreated, death from liver failure, recurrent bleeding, liver cancer or neurological crisis frequently occurs within the first 10 years of life.

The aim of the report is to examine three key questions relating to the effectiveness and appropriateness of newborn screening using tandem mass spectrometry (TMS) for TYR1. Specific questions for the review are shown below:

1. What is the incidence of TYR1 in the UK? (NSC criterion 1: The epidemiology, incidence, prevalence and natural history of the condition should be understood)
2. What is the test accuracy (sensitivity, specificity, and predictive values applicable to UK prevalence) of succinylacetone measurement in dried blood spots (DBS) using tandem mass spectrometry for TYR1 screening? (NSC criterion 4: There should be a simple, safe, precise and validated screening test.)
3. Does early treatment with Nitisinone (NTBC, Orfadin®) following screening provide better long-term outcomes than later treatment with Nitisinone after the presentation of symptoms? (NSC criterion 9: There should be an effective intervention for patients identified through screening, with evidence that intervention at a pre-symptomatic phase leads to better outcomes for the screened individual compared with usual care.)

Methods

We used a rapid review approach. We undertook separate literature searches for each of the three key questions. For key question 1 searches started from August 2012 (the cut-off date of the initial 2014 review[2]). There were no date limits for the searches for key question 2 and 3. Searches were undertaken on 14th September 2015 for key questions 2 (screening test) and 3 (treatment) and on 22nd September 2015 for key question 1 (incidence). Searches were conducted in Web of Science (Core Collection), Medline (Ovid), Medline In-Process & Other Non-Indexed Citations (Ovid), Embase (Ovid) and the Cochrane Library. Reference lists of all included articles were screened.

One reviewer screened the titles and abstracts of all records identified by the searches, and decided whether to include each one based on the full text of all articles deemed potentially relevant. One reviewer extracted the information needed using a standard data extraction form. Formal quality assessment was not undertaken for key question 1. To assess the quality of the studies for questions 2 and 3 standard quality assessment tools were used (Question 2: untailored QUADAS-2; Question 3: Effective Public Health Practice Project [EPHPP] quality assessment tool for quantitative studies). A second reviewer independently repeated each stage for a random sample of 20% of records (data extraction was checked by a second reviewer). Disagreements were resolved by consensus or through discussion with a third reviewer. Study design, treatment, population, and outcome characteristics were summarised in text and tables. We did not combine study results using a meta-analysis.

Results

Key question 1 (Incidence of TYR1 in the UK)

We did not find any additional studies which investigated the incidence of TYR1 in the UK or in Western-European countries (published after August 2012). Four included studies reported incidence estimates for Tunisia (2 studies), United Arab Emirates and Singapore, but results cannot be applied to the UK newborn population as ethnicity and consanguinity rates are not comparable. Three further studies identified by additional searches gave incidence estimates of TYR1 for Bahrain, Lebanon and all 51 U.S. states. The only data on TYR1 in the UK (published in the 1990s) suggest an incidence of about 1 in 55,000 in the West Midlands, with an approximately 100-fold higher incidence among Pakistanis (1:2,628) than among North-Western Europeans (1:302,665).

NSC criterion 1: Not met

Key question 2 (Screening test using succinylacetone)

Ten articles were included in this review; four case-control studies used stored DBS samples from confirmed TYR1 patients and healthy controls, and six studies reported results from prospective newborn screening programmes. Extraction methods and SUAC cut-offs employed differed between the studies.

Risk of bias was considered high in two or more domains in four of the 10 identified studies (40%) and in one domain in the remaining six studies (60%). No study was judged as low or unclear risk of

bias in all four domains. Study flow was assessed as at high risk of bias in 8/10 studies, (80% high risk). The major problems were that screen-positives and screen-negatives (or cases and controls) did not receive the same reference standard to verify their TYR1 status, that follow-up of screen-negatives was not defined or not conducted, and that losses to follow-up were not reported. There were only a few cases of TYR1 in the prospective studies, and case-control studies produce inflated estimates of accuracy.

There were significant concerns regarding applicability of the research identified to the UK screening population in six out of the 10 (60%) included studies. This is because the TYR1 incidence was higher than expected in the UK population and/or screened DBS samples were collected before 5 days or after 8 days of life.

The determination of sensitivity, specificity, and negative predictive value were not possible from the included prospective screening studies. Positive predictive values (PPV) from four prospective screening studies using SUAC as primary marker was 100% in three studies (6 true positive cases out of 717,501 people screened) and 67% in one study (2 true positive cases and 1 false positive case out of ~500,000 people screened). PPV could not be calculated in two studies. There were very wide confidence intervals due to the small number of cases. As none of the studies were conducted in the UK, the applicability of the PPV to the birth prevalence of TYR1 in the UK is reduced.

NSC criterion 4: Not met.

Key question 3 (Early vs. late Nitisinone treatment)

The results of two prospective cohorts (one UK, one Canadian) and one international survey examining the outcomes for people with TYR1 treated with nitisinone at different start ages were reported in six papers. One reported prospective and retrospective data collection, while all others reported retrospective only. The number of included TYR1 patients per study ranged from 17 from a single centre in the UK, to 168 from a survey of 21 centres in Europe, Turkey and Israel. The methodological quality was moderate (one weak rating) in three papers and weak (two or more weak ratings) in the three other papers. There was high risk of selection bias in two papers and all six papers had a high potential for confounding as important factors (i.e. pre-existing health problems, presenting form of TYR1, compliance with treatment, co-treatment) were not considered in study design or analysis. There was also a wide overlap of included TYR1 patients between papers.

There is evidence from these papers suggesting that Nitisinone is an effective treatment for TYR1. However, evidence on the benefit of early versus late treatment is unclear. There was some evidence that early treatment with Nitisinone and diet within the first two months of life might be associated with a reduced mortality rate and reduced need for liver transplantation compared to later treatment initiation. However, these potential effects have not been reported consistently between studies and it is not clear whether outcomes in groups given early administration of Nitisinone are due to treatment effectiveness, differences in the spectrum of disease, methodological biases, chance findings from small samples, or other confounding factors. Furthermore, there were applicability concerns because many of the late treatment group were cases diagnosed before the introduction of Nitisinone, so they would have been treated significantly later in life than is the case in the UK currently, where Nitisinone treatment starts soon after diagnosis.

NSC criterion 9: Not met.

Conclusions and implications for policy

More research is needed to evaluate the incidence of TYR1 in the UK, as well as to examine the value of current neonatal screening programmes for PKU and cascade testing of children with affected siblings for detecting TYR1.

A research project using tandem mass spectrometry measurement of SUAC from dried blood spots with follow-up of screen-negatives for at least two years would considerably strengthen the test performance data; this could be achieved through follow-up of one of the existing cohorts.

For the treatment, further investigation is needed regarding whether the TYR1 cases detected by screening represent the same spectrum of disease as those detected symptomatically and whether there is certainty that all screen-detected babies would become symptomatic in the absence of screening. Furthermore, research is needed to investigate whether improved outcomes with early administration of Nitisinone are due to the effectiveness of the drug, differences in the spectrum of disease or other confounding factors. In addition clarification is needed as to whether improved outcomes with early detection are applicable to the UK question (i.e. whether the early detected cases are sufficiently similar to SUAC screen-detected cases in a potential UK programme and whether the late detected cases are sufficiently similar to symptomatically detected tyrosinemia in the UK).

1. Introduction

Tyrosinaemia type 1 (TYR1, OMIM 276700) is a rare autosomal recessive disorder of amino acid metabolism that is caused by a defect in the final enzyme of the pathway of tyrosine degradation, namely fumarylacetoacetate hydrolase (FAH, also known as fumarylacetoacetase). About one person in 100,000 is affected with TYR1 globally, but incidence is more common in some regions, notably in Québec.[1] Deficiency of FAH causes an accumulation of tyrosine and toxic metabolites succinylacetone (SUAC), maleylacetoacetate and fumarylacetoacetate (Figure 1). TYR1 mainly affects the liver, kidneys and peripheral nerves. Symptoms may start during the first few months of life in the acute form of TYR1. The chronic form results in more gradual liver disease; hepatocellular carcinoma is a frequent complication. Without treatment, death from liver failure, recurrent bleeding, neurological crisis or hepatocellular carcinoma frequently occurs before the age of 10 years.[3]

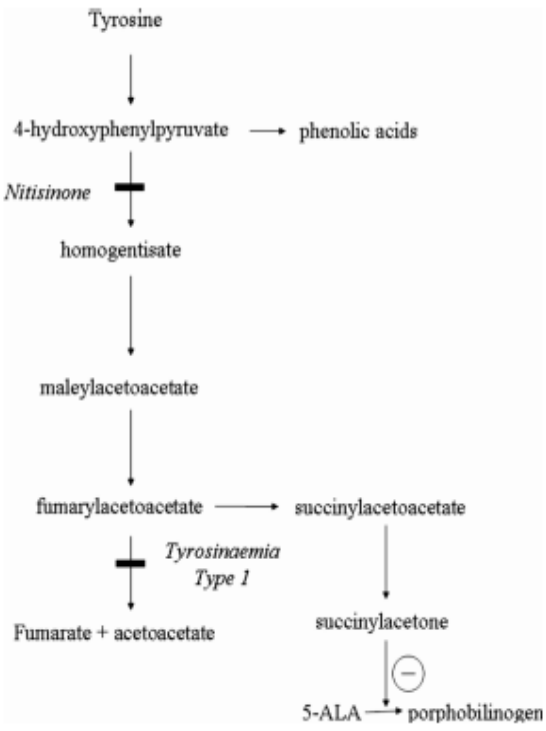

Figure 1. Tyrosine degradation pathway

TYR1 is caused by a defect in fumarylacetoacetase. Succinylacetone is a potent inhibitor of 5-aminolaevulinate dehydratase. Nitisinone inhibits 4-hydroxyphenylpyruvate dioxygenase reducing flux through the pathway.[1]

TYR1 is treated with Nitisinone (2-[2-nitro-4-trifluoromethylbenzoyl]-1,3-cyclohexanedione, NTBC, Orfadin®) accompanied by a protein-restricted diet (usually low in phenylalanine, methionine and tyrosine). Nitisinone inhibits 4-hydroxyphenylpyruvate dioxygenase, an enzyme that is upstream of FAH in the tyrosine degradation pathway, and so reduces the formation of toxic metabolites (see Figure 1). At the present time, liver transplantation remains the only effective means of establishing normal enzyme activity and treating the disease in patients with hepatocellular carcinoma or with failure to respond to Nitisinone treatment.

The current NHS Newborn Blood Spot (NBS) Screening Programme recommends that all newborn babies in England are screened for the following nine conditions: sickle cell disease (SCD), cystic fibrosis (CF), congenital hypothyroidism (CHT), phenylketonuria (PKU), medium-chain acyl-CoA dehydrogenase deficiency (MCADD), maple syrup urine disease (MSUD), isovaleric acidaemia (IVA), glutaric aciduria type 1 (GA1), and homocystinuria (HCU). The same nine conditions are part of the NBS screening programme in Wales. In Scotland, newborns are screened for SCD, CF, CHT, PKU and MCADD. TYR1 is not included in the current NBS Screening Programme in England, Scotland, Wales or Northern Ireland. For NBS screening, a heel-prick blood sample is collected on a standard newborn screening collection card from babies on day 5 (in exceptional circumstances between day 5 and day 8) and sent to the regional newborn screening laboratory. Dried blood spots are analysed using tandem mass spectrometry (TMS), alongside a variety of other techniques.

Basis for current recommendation

The most recent UK NSC update review of screening for TYR1 was signed off at the March 2015 UK NSC meeting.[2] The review contained literature from relevant publications between 2004 and August 2012. The review suggested additional work focussing on specific questions to assess the possibility of screening for this condition. A public consultation highlighted key issues which might provide a focus for further consideration of the evidence base relating to screening for this condition.

Current update review and approach taken

This review is structured to explore the following:

- Inclusion of studies published since August 2012, which will increase the understanding of TYR1 incidence.
- Short-term and long-term outcomes of Nitisinone treatment.
- Focus on studies using succinylacetone (SUAC) as the single TMS screening marker for TYR1.

This review will inform discussion and decision making about further work on the viability of TYR1 screening, for example primary research, cost effectiveness analysis, disease modelling.

2. Aims

The aim of the evidence review is to examine three key questions relating to the effectiveness and appropriateness of newborn screening of SUAC levels in DBS using TMS for TYR1. The key questions for this project are shown in the table below:

Key questions	**NSC criterion**
1. What is the incidence of TYR1 in the UK?	1. The condition should be an important health problem as judged by its frequency and/or severity. The epidemiology, incidence, prevalence and natural history of the condition should be understood, including development from latent to declared disease and/or there should be robust evidence about the association between the risk or disease marker and serious or treatable disease.
2. What is the test accuracy (sensitivity, specificity, and predictive values applicable to UK prevalence) of SUAC measurement in dried blood spots using TMS for TYR1 screening?	4. There should be a simple, safe, precise and validated screening test.
3. Does early treatment with Nitisinone (NTBC, Orfadin®) following screening provide better long-term outcomes than later treatment with Nitisinone after the presentation of symptoms?	9. There should be an effective intervention for patients identified through screening, with evidence that intervention at a pre-symptomatic phase leads to better outcomes for the screened individual compared with usual care. Evidence relating to wider benefits of screening, for example those relating to family members, should be taken into account where available. However, where there is no prospect of benefit for the individual screened then the screening programme should not be further considered.

3. Methods

We used a rapid evidence assessment approach (REA). The UK NSC has produced a set of requirements for evidence summaries, or REAs, for use in its evidence review process which provided the framework for the conduct of the review. We also sought guidance from commissioners and area experts.

3.1. Identification and selection of studies

Separate literature searches were performed for each of the three key questions. Copies of the search strategies used in the major databases are provided in Appendix 1 A-C. For key question 1, searches started from August 2012 (the cut-off date of the initial 2014 review.[2]). No date limits were applied to the searches for key question 2 and 3. Searches were undertaken on 14th September 2015 for key questions 2 (screening test) and 3 (treatment) and on 22nd September 2015 for key question 1 (incidence). Searches were conducted in Web of Science (Core Collection), Medline (Ovid), Medline In-Process & Other Non-Indexed Citations (Ovid), Embase (Ovid) and the Cochrane Library. Reference lists of all included articles were screened. Table 1 shows the inclusion and exclusion criteria used for the three key questions.

Table 1. Inclusion and exclusion criteria for the three key questions

Key question	Inclusion criteria							Exclusion criteria
	Population	**Target condition**	**Intervention**	**Reference Standard**	**Comparator**	**Outcome**	**Study type**	
1) Incidence	All general populations not at high risk of inborn errors of metabolism.	Tyrosinaemia type 1	None	None	None	Incidence of TYR1	Any systematic review, cross-sectional study or cohort study ideally taken over at least five years	Papers published before 2012, non-human studies, papers not available in the English language, letters, editorials and communications, grey literature and conference abstracts.
2) Screening test	Neonatal or newborn infants	Tyrosinaemia type 1	The index test is newborn screening for TYR1 using TMS measurement of SUAC in dried blood spots	Urine testing for SUAC and/or subsequent clinical detection of TYR1	None	Sensitivity, specificity, predictive values	Cross-sectional test accuracy studies, case-control studies and cohort studies	Non-human studies, papers not available in the English language, letters, editorials and communications, grey literature and conference abstracts.
3) Treatment	Patients with tyrosinaemia type 1	Tyrosinaemia type 1	Early treatment with Nitisinone (NTBC, Orfadin®) following screening (universal newborn screening, cascade testing or incidental detection)	None	Later treatment with Nitisinone following presentation of symptoms	Any outcome of treatment	Any study design in humans	Non-human studies, papers not available in the English language, letters, editorials and communications, grey literature and conference abstracts.

SUAC, succinylacetone; TMS, tandem mass spectrometry; TYR1, tyrosinaemia type 1.

3.2. Review strategy

One reviewer screened the titles and abstracts of all records identified by the searches. A second reviewer independently screened a random sample of 20% of titles and abstracts. Disagreement was resolved through consensus or discussion with a third reviewer. Full copies of all studies deemed potentially relevant were assessed for inclusion by one reviewer. The second reviewer independently screened a random sample of 20%; disagreements were resolved by consensus or discussion with a third reviewer. Records rejected at full text stage and reasons for their exclusion were documented (Appendix 2, Appendix 3, and Appendix 4).

3.3. Data extraction strategy

An electronic, piloted data extraction form was used to extract data by one reviewer. A second reviewer checked a random sample of 20% of articles as a quality control measure. Disagreements were resolved by consensus or discussion with a third reviewer. An example of a data extraction sheet is provided in Appendix 5.

3.4. Assessment of study quality

Formal quality assessment was not undertaken for key question 1. For the quality appraisal of questions 2 and 3 standard quality assessment tools were used (Question 2: untailored QUADAS-2[4]; Question 3: Effective Public Health Practice Project (EPHPP) quality assessment tool for quantitative studies[5]; see Appendix 6A & B). Quality assessment was undertaken by one reviewer; a second reviewer independently appraised the quality of a random sample of 20% of studies. Disagreements were resolved by consensus or through discussion with a third reviewer.

3.5. Methods of analysis/synthesis

Study design, treatment, population, and outcome characteristics were summarised in text and tables. TYR1 incidence data has been split into screened versus unscreened populations. Pooling study results by meta-analysis was not performed.

4. Results: Appraisal against UK NSC Criteria

The full list of the UK NSC criteria is available in Appendix 7.

4.1. Key question 1 (Incidence)

What is the incidence of TYR1 in the UK?

This relates to NSC criterion 1:

'The condition should be an important health problem as judged by its frequency and/or severity. The epidemiology, incidence, prevalence and natural history of the condition should be understood, including development from latent to declared disease and/or there should be robust evidence about the association between the risk or disease marker and serious or treatable disease.'

Description of the evidence

Figure 2 provides the PRISMA flow diagram for the incidence review. Our electronic searches identified 121 unique records. Another three references were identified by additional searches. Seventeen full text articles were assessed of which 10 articles were subsequently excluded using the pre-defined inclusion / exclusion criteria (see Appendix 2 for excluded studies with reason). This left 7 articles that met the inclusion criteria and were included in the narrative synthesis.

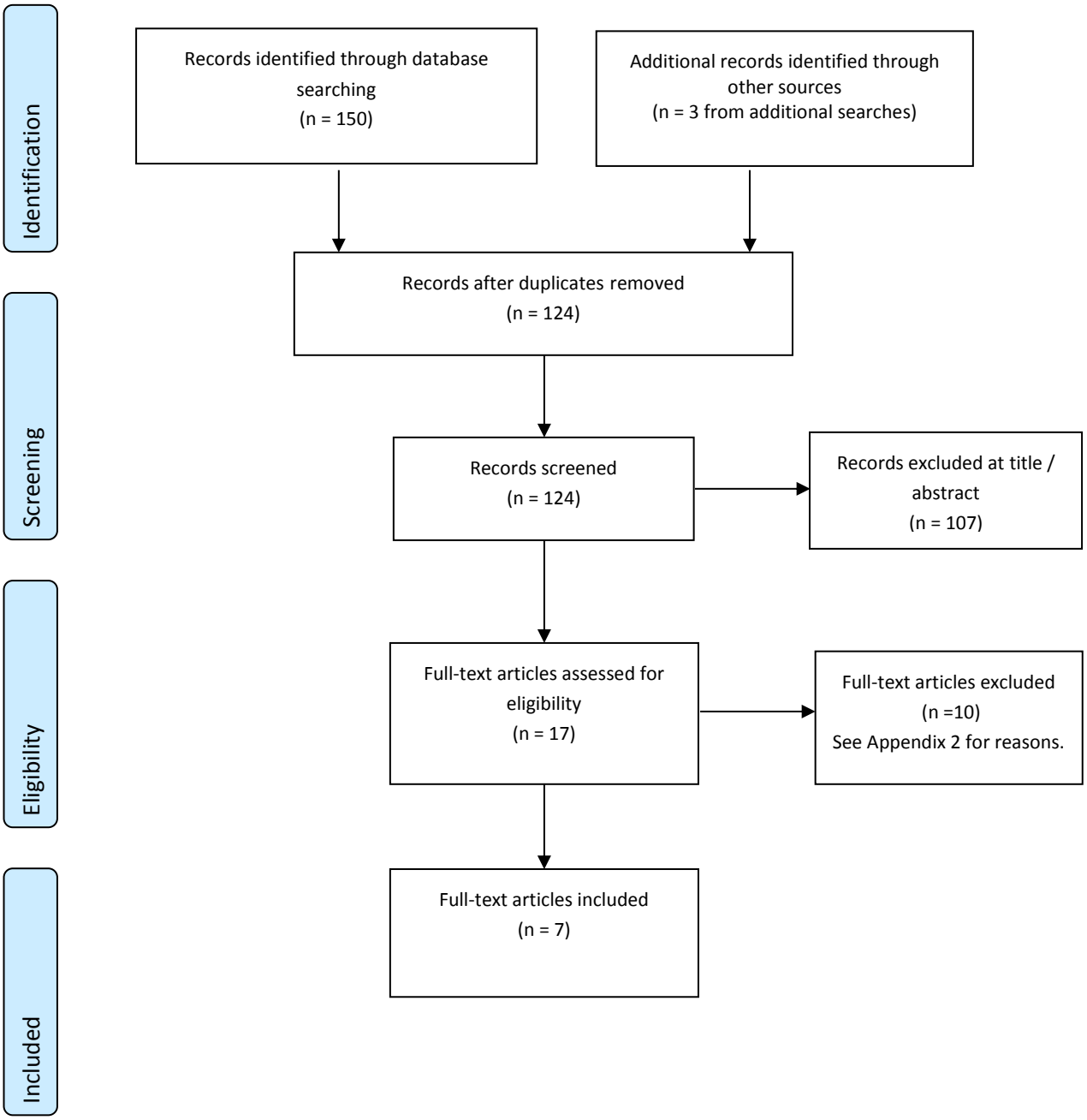

Figure 2. PRISMA Flow Diagram: Incidence of TYR1

Analysis of the evidence

The previous review by Bazian Ltd. (2014)[2] concluded that *"…the incidence of tyrosinaemia type I varied widely between reports from different countries, from more than one case per 31,000 births to less than one case per 944,000 births. Where a screened cohort was compared to a contemporaneous or historical unscreened cohort in the same country, one study found that the incidence of tyrosinaemia type I was similar. Another study found the incidence increased, despite the fact that the cases were missed by screening. It should be noted that individual cases of the disease have a big impact on calculated incidences as tyrosinaemia type I is a rare disease."* Incidences from studies published between 2004 and 2012 as identified by the previous review are shown in Table 2.

No studies published after August 2012 (the cut-off date of the previous review by Bazian Ltd[2]) which investigated the incidence of TYR1 in the UK or Western-European countries were identified. Four studies reported incidence estimates for Tunisia (2 studies[6; 7]), United Arab Emirates[8], and Singapore[9] (Table 3). Additional searches led to the identification of three further studies reporting incidence estimates for Bahrain[10], Lebanon[11] and all 51 U.S. states.[12] In Tunisia, the incidence of TYR1 was 1:14,804 in a cohort of unscreened children born between 1987 and 2009.[6] Exactly the same incidence (1:14,804) was found in a Tunisian cohort of unscreened children born between 1988 and 2013, [7]. These findings strongly suggest that these are two reports of the same study. In the United Arab Emirates, the incidence of TYR1 was in the range of 2.2-4.9:100,000 live births (corresponding to approximately 1:20,408 -1:45,455) in an unscreened cohort.[8] In Singapore, Lim et al. (2014) found no cases of TYR1 when screening 177,267 newborns born between July 2006 and April 2014 in the Singapore Newborn Screening Programme.[9] No patient was diagnosed with TYR1 in about 140,000 unscreened newborns during the same period. In Bahrain, the incidence of TYR1 was 1:33,282 in a cohort of 66,565 unscreened children born between January 2008 and December 2011.[10] In Lebanon, among 126,000 samples screened between 2007 and 2013, one case of TYR1 was confirmed, yielding an incidence of 1:126,000.[11] Therrell et al. (2014) reported 10-year incidence data (January 2001 to December 2010) for selected Inborn errors of metabolism (IEM) detected through newborn screening in the U.S. from 51 national partnering programs (50 states and the District of Columbia).[12] Screening for TYR1 was first introduced in seven states in 2001 and the last three U.S. states initiated screening in 2009; the time period of TYR1 newborn screening ranged from 12 months to 10 years. Screening test and case definition/confirmation of disease differed between the states and were not reported. The total incidence of TYR1 in an estimated 24,521,197 screened newborns across all 51 states was reported as 1:781,144 (1:681,144 when

calculated by reviewers) ranging from 1:29,416 in the District of Columbia to 1:1,417,732 in Georgia and no TYR1 cases detected by newborn screening in 31 of 51 states (see Appendix 8 for incidence data for individual U.S. states).

Discussion: Question 1

The reported incidences of TYR1 ranged from 1:14,804 in an unscreened population in Tunisia to 1:1,417,732 detected through newborn screening in Georgia (USA). The underlying population size was not reported in two studies.[6; 8] Increased disease frequency of TYR1 can be due to parental consanguinity and/or an increased rate of genetic mutation in certain ethnic groups.[13] In Tunisia, a first-degree consanguinity rate of 65% was found in 69 TYR1 cases. As ethnic groups and consanguinity rate in North Africa, the Middle East, and Asia are different to those in the UK, the observed incidences of TYR1 in Tunisia, United Arab Emirates, Bahrain, Lebanon or Singapore cannot be applied to the UK population. Incidence of TYR1 in screened newborns across all 51 U.S. states was estimated to be 1:781,144 (1:681,144 when calculated by reviewers) but screening test, case definition and duration of TYR1 screening differed across states.[12] The reported incidence data were based on estimated numbers of births from a voluntary national data system that depends on unfunded cooperation and collaboration among the 51 national partnering programmes and might be prone to inaccuracies. Therrell and co-workers[12] assumed 100% screening coverage and used births reported by vital record managers and not the NBS programme as population base for their incidence calculations; false-negatives and TYR1 cases in unscreened babies were therefore not included in the incidence estimate and may be underestimated.

The only data on the incidence of TYR1 in the UK are based on two retrospective cohort studies in the West Midlands Regional Health Authority (RHA) (the largest health region in England and Wales covering the counties of the West Midlands, Hereford, Worcester, Shropshire, Staffordshire, and Warwickshire).[13; 14] The incidence of TYR1 in 670,718 children born between January 1985 and March 1994 in the West Midlands RHA was 1 in 55,893.[14] Ten of 12 patients (83%) with TYR1 were of Asian ethnicity. A second study in 707,720 children born between April 1981 and April 1991 in the West Midlands found an incidence of TYR1 of 1 in 54,440. Incidence of TYR1 was approximately 100-fold greater among Pakistanis (1:2,628) than among North-Western Europeans (1:302,665, $p<0.001$).[13]

Additional literature searches for key question 2 (screening test) using broader search terms i.e. for the condition ('inborn errors of metabolism' instead of 'TYR1' only) identified three more articles reporting incidence estimates of TYR1. Therefore, a full systematic review may yield more results,

but due to low transferability of results between countries may not answer the question. A new research study investigating UK incidence is required.

Summary

Criterion 1: Not met.

The previous review by Bazian published in 2014[2] concluded: '*No studies were identified that reported the incidence or prevalence of tyrosinaemia type I in the UK. The natural history of the condition is well established. Tyrosinaemia type I can be detected through MS/MS analysis of newborn blood spots. The concentrations of tyrosine and succinylacetone can be used to screen for tyrosinaemia type I. Succinylacetone is a more sensitive and specific marker of tyrosinaemia type I as tyrosine levels can be elevated due to transient tyrosinaemia, liver disease, or tyrosinaemia type II or type III, and because some infants with tyrosinaemia type I may have normal blood concentrations of tyrosine when the screening sample is taken. The duration of the latent asymptomatic period varies, and no marker has been identified that can discriminate between the early and late onset forms. It has been reported that infants with tyrosinaemia type I can develop symptoms before the results of a newborn dried blood spot screen are available*.'

This update review did not identify any studies published since 2012 which reported incidence of TYR1 in the UK or North-Western Europe. Two studies published in 1996 and 1998 report incidence of TYR1 in the West Midlands region of 1:55,893 and 1:54,440, respectively, with an approximately 100-fold higher incidence among Pakistanis (1:2,628) than among North-Western Europeans (1:302,665). Data from the mid to late 1990s suggested the incidence of TYR1 in the West Midlands (excluding Birmingham) was 1:105,037. An update on the incidence of TYR1 in the UK is required.

Table 2. Estimates of the incidence of tyrosinaemia type 1 published between 2004 and 2012 [identified by Bazian (2014)[2]]

Reference	Country and time period	Ethnicity	Screening programme in operation?	Population size	Number of identified cases	Incidence (per number of live births)	Notes
Bliksrud 2012[15]	Norway 1 January 1991 to 31 December 2010 (20 years)	NR	No	1,181,636	15.8 (14 diagnosed, 1.8 undiagnosed)	1:74,800	A point estimate of the undiagnosed patients was made based on our patients' age at the time of diagnosis.
Lund 2012[16] Screened	Denmark, the Faroe Islands and Greenland 2 February 2009 to 31 March 2011 (2 years)	NR	Yes [PerkinElmer NeoBase non-derivatized MS/MS kit™ (3040-0010), SUAC > 2.1 U]	140,565	1	1:140,565	
Unscreened historic cohort	Denmark, the Faroe Islands and Greenland 1 January 1992 to 31 December 2001 (10 years)	NR	No	674,754	NR *(6)*	1:112,459	
Couce 2011[17]	Galicia (Spain) July 2000 and July 2010 (10 years)	NR	Yes [Tyr > 175 µM]	210,165	2 by screening; 1 missed by screening	1:105,082 *1:70,055 when 1 FN is included*	1 case of TYR1 given a false negative result.
Lindner 2011[18]	3 states in South-Western Germany 1 January 1999 to 30 April 2005 (6 years)	NR	Yes [Tyr > 200 µmol/l, spectrophotometric microassay for ALAD as 2nd-tier test, from [19]]	583,553	2 (plus 1 screen-positive unconfirmed due to early death)	1:291,777 *1:194,518 when unconfirmed case is TP)*	1 screen-positive case unconfirmed due to early death.
Morrissey 2011[20]	New York state (USA) December 2007 to December 2009 (2 years)	NR	Yes [SUAC > 3.0 µmol/l for retest in duplicate, Average (initial and retest) 3.0 – 5.0 µmol/l: request repeat DBS; Average (initial and retest) ≥ 5.0 µmol/l: Immediate referral]	~500,000	2	*~1:250,000*	No exact number of screened samples.
Kasper 2010[21]	Austria April 2002 to December 2009 (8 years)	NR	Yes [1) Tyrosine 2) ALAD activity as 2nd-tier.	622,489	5	1:124,498	

Reference	Country and time period	Ethnicity	Screening programme in operation?	Population size	Number of identified cases	Incidence (per number of live births)	Notes
Wilcken 2009[22] Screened	Australia 1 April 1998 to 31 March 2002 (4 years)	NR	Yes (depending on area) [Tyrosine as marker]	461,500	2 missed by screening	*1:230,750*	2 missed by screening using tyrosine as marker analyte.
Unscreened	Australia 1 April 1998 to 31 March 2002 (4 years)	NR	No	533,400	2	*1:266,700*	
Unscreened historical cohort	Australia 1 April 1994 to 31 March 1998 (4 years)	NR	No	1,017800	1	*1:1,017,800*	
Total unscreened	Australia 1 April 1994 to 31 March 2002 (8 years)	NR	No	1,551,200	3	*1:517,067*	
La Marca 2008[23]	Tuscany (Italy) January 2002 to 2008 (6 years)	NR	Yes [01/2002-12/2006: Tyr > 200 µmol/l; From 01/2007: SUAC > 2 µmol/l, Tyr > 250 µmol/l as secondary marker).]	160,000	1 missed by screening	*1:160,000 when 1 FN case included*	1 TYR1 case missed by screening using tyrosine as marker analyte.
Masurel-Paulet 2008[24]	France 1990 - NR	NR	No	NR	74	<1:200,000	
Feuchtbaum 2006[25] Screened	California (USA) January 2002 to June 2003 (18 months)	NR	Yes [NeoGram amino acids and acylcarnitines derivatized MS/MS reagent kit (PerkinElmer)].	353,894	0	NA	
Unscreened	California (USA) January 2002 to June 2003 (18 months)	NR	No (participation depended on cooperation of maternity hospitals throughout the state)	401,779	0	NA	

Reference	Country and time period	Ethnicity	Screening programme in operation?	Population size	Number of identified cases	Incidence (per number of live births)	Notes
Frazier 2006[26]	North Carolina (USA) July 1997 to July 2005 (8 years)	Between 1999 and 2004: 73% Caucasian, 23% African American, 2.5% Asian, 1.5% Native American. Hispanic infants about 12% of all newborns.	Yes [Tyr as marker, cut-off modified over time, Since 1 January 2003: Borderline: > 500 µmol/l; diagnostic: > 900 µmol/l].	944,078	1 missed by screening	*1:944,078 when 1 FN included*	1 case missed by screening using tyrosine as marker analyte.
Sander 2006[27]	Germany 16 weeks	NR	Yes [SUAC > 10 µmol/l}	61,344	2	*1:30,672*	The diagnostic sensitivity has not been evaluated with follow-up of all 61,344 newborns for TYR1.
Comeau 2004[28]	New England Newborn Screening Programme (USA) January 1999 to January 2003 (4 years)	NR	Yes [Tyr > 442 µmol/l, Tyr/Phe > 6, from [29]]	318,535	0	NA	

ALAD, 5-aminolevulinic acid dehydratase; FN, false negative; IEM, inborn errors of metabolism; NA, not applicable; NBS, newborn screening; NR, not reported; Phe, phenylalanine; SUAC, succinylacetone; TN, true negative; TP, true positive; Tyr, tyrosine; TYR1, tyrosinaemia type 1.

Numbers in italics were calculated by reviewers.

Table 3. Estimates of the incidence of tyrosinaemia type 1 (published since 2012)

Reference	Country and time period	Ethnicity	Screening programme in operation?	Population size	Number of identified cases	Incidence (per number of live births)	Notes
Al-Shamsi 2014[8]	United Arab Emirates January 1995 to December 2011 (16 years)	NR (Emiratis)	No	NR	2	*1:20,408 - 1:45,455*	22 Emirati patients with IEM managed at the Latifa Hospital in Dubai not included. Exact incidence of TYR1 NR, only range for 9 conditions given.
*Golbahar 2013[10]	Bahrain 1 January 2008 to 31 December 2011 (4 years)	NR (Bahrain)	No	66,565	2	1:33,282	
*Khneisser 2015[11]	Lebanon	NR (Lebanon)	Yes [TMS based screening test]	126,000	1	*1:126,000*	Screening test NR
Lim 2014[9] Screened	Singapore July 2006 to April 2014 (8 years)	NR (Singapore: 74.2% Chinese, 13.4% Malays, 9.9% Indians, 3.2% other)	Yes but voluntary [July 2006-December 2010: Tyr as marker; From January 2011: 2nd-tier SUAC ≥ 5.0 µmol/l]	177,267	0	NA	
Unscreened	Singapore July 2006 to April 2014 (8 years)	NR (Singapore: see above)	No	~140,000	0	NA	
Hadj-Taieb 2012[6] Nasrallah 2015[7]	Tunisia 1987 to 2009 inclusive (23 years) and 1988 to 2013 (25 years)	NR (Tunisia)	No	*25*166,000= 4,150,000*	69† (61 in earlier study)	1:14,804	Birth prevalence was calculated using Hardy-Weinberg formula with coefficient of consanguinity = 0.0129. † 3 prenatally diagnosed TYR1 cases terminated and possibly not included in this number.
*Therrell 2014[12]	51 U.S. states January 1, 2001 to December 31, 2010 (up to 10 years)	NR (United States)	Yes [Screening test NR, possibly differences between 51 U.S. states]	24,521,197	36	1:781,144 *1:681,144*	*Incidence 1:681,144* (error in publication?); Individual states introduced TYR1 newborn screening between 2001 and 2009 (see Appendix 8); TYR1 incidence for individual U.S. states shown in Appendix 8.

IEM, inborn errors of metabolism; NA, not applicable; NBS, newborn screening; NR, not reported; Phe, phenylalanine; SUAC, succinylacetone; Tyr, tyrosine; TMS, tandem mass spectrometry; TYR1, tyrosinaemia type 1.

* References retrieved from additional searches

Numbers in italics were calculated by reviewers.

4.2. Key question 2 (Screening test)

What is the test accuracy (sensitivity, specificity, and predictive values applicable to UK prevalence) of SUAC measurement in dried blood spots using TMS for TYR1 screening?

This relates to NSC criterion 4:

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

Description of the evidence

Figure 3 provides the PRISMA flow diagram for the screening test review. Our electronic searches identified 310 unique records; one additional reference was identified by screening reference lists of included studies. Twenty full text articles were assessed, of which 10 were subsequently excluded using the pre-defined inclusion / exclusion criteria (see Appendix 3 for excluded studies with reason). This left 10 articles which met the inclusion criteria and were included in the narrative synthesis.

Characteristics of included studies

Included studies are summarised in Table 4. There were ten studies: six studies reported results from newborn screening programmes,[16; 20; 27; 30; 31; 32] ranging from 4,683 samples collected over 1 month[31] to 515,592 samples collected over 4 years 1 month,[32] and four case-control studies used stored DBS samples from confirmed TYR1 patients and healthy controls.[33; 34; 35; 36]

The TMS methods employed differed between the studies: two studies used commercially available TMS assays[16; 31] while all others (n=8) used non-kit methods with derivatisation of SUAC to its hydrazone.[20; 27; 30; 32; 33; 34; 35; 36] TMS analysis of SUAC-hydrazone[20; 33); (27; 32; 34] or SUAC-hydrazone butyl ester [30; 31; 35; 36] was performed; the TMS methodology used by Lund et al. (2012)[16] was not reported. SUAC cut-off values used in the 10 studies ranged from 1.29 µmol/l[31] to 10 µmol/l.[27] A detailed description of the study design and the TMS methodology used can be found in Appendix 9.

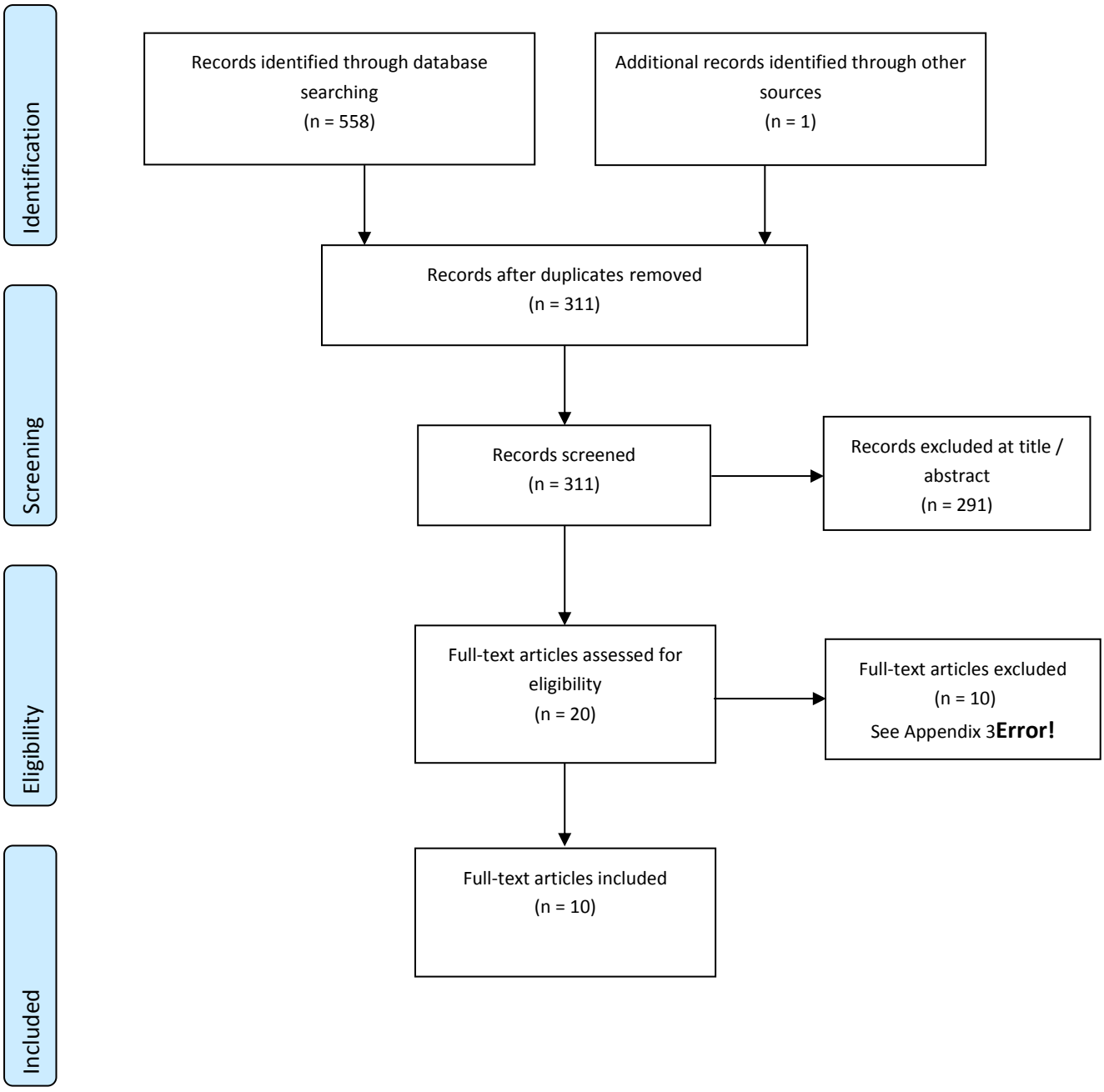

Figure 3. PRISMA Flow Diagram: Newborn TMS screening for TYR1 using SUAC

Methodological quality of included studies

The methodological quality of the 10 included studies, assessed by untailored QUADAS-2[4] is summarised in Figure 4, Figure 5, and Appendix 10. Risk of bias was considered high in two or more domains in four of 10 studies (40%) and in one domain in the remaining six studies (60%). No study was judged as low or unclear risk of bias in all four domains.[1] Figure 4 shows that study flow was the area with the greatest risk of bias (8/10, 80% high risk); no study scored a low risk of bias in this domain. The major problem was that screen-positives and screen-negatives (or cases and controls) did not receive the same reference standard to verify their TYR1 status. Another issue was incomplete or unclear reporting, particularly of the conduct of the index test (i.e. how cut-off value was determined) and the reference standard, which is reflected in high proportions (6 [60%] and 9 [90%] of 10 studies) scoring an unclear risk of bias in these two domains, respectively. The method and duration of clinical follow-up in screen-negatives was not described in any prospective screening study (n=6), it is not clear whether any study followed up screen negatives at all. Three of four case-control studies did not describe the reference standard used.

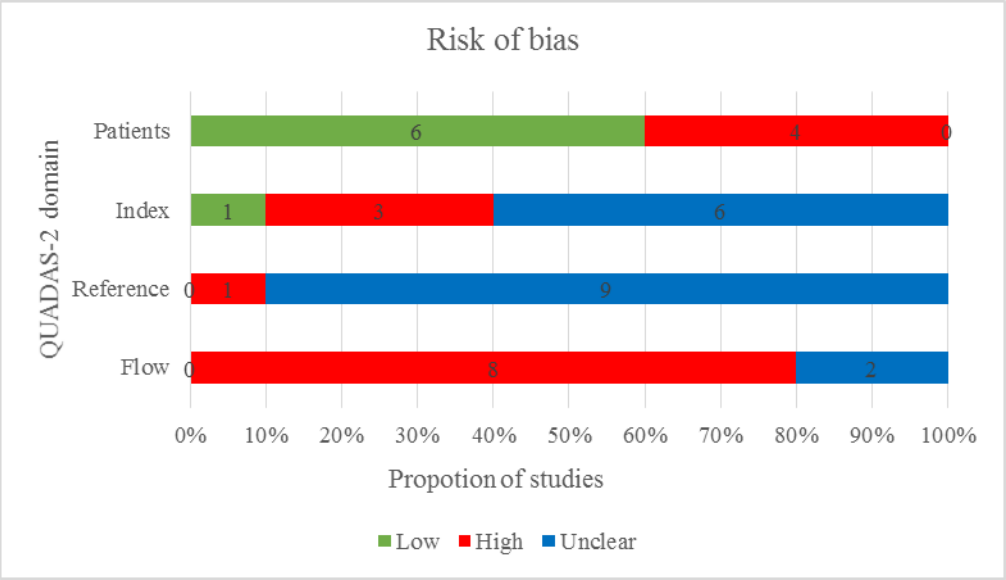

Figure 4. Risk of bias in included screening studies (untailored QUADAS-2[4])

There were significant concerns regarding applicability of the research identified to the UK screening population in six out of the 10 (60%) included studies (see Figure 5). This is because the TYR1 incidence in one prospective NBS study was higher (≥1:30,672) than expected in the UK

[1] Methodological quality using adjusted QUADAS-2 with guidance notes is given in Appendix 11 with differences highlighted in Appendix 13.

population,[27] five studies collected DBS samples for screening earlier than 5 days of life,[16; 27; 30; 33; 36] and two studies included screening samples that were collected after one month of age in some children.[32; 36]

Concerns regarding the applicability of the index test to the situation in the UK were classified as unclear in two studies using an indeterminate SUAC range and repeat DBS request.[20; 32]

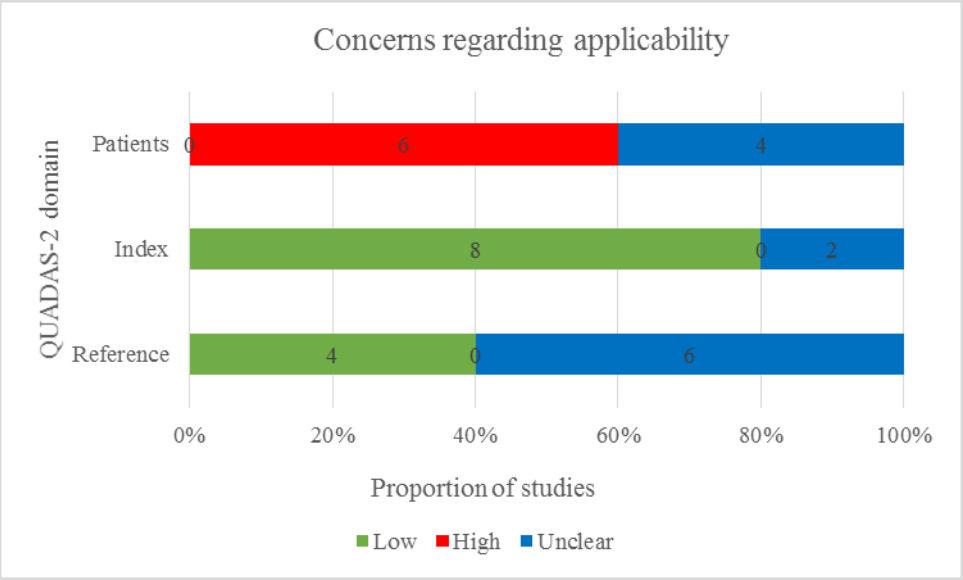

Figure 5. Concerns regarding applicability in included studies (untailored QUADAS-2[4])

Analysis of the evidence

Results from four included case-control studies showed a clear discrimination of SUAC levels between affected and unaffected newborns. The determination of sensitivity, specificity, and negative predictive value was not possible from the included prospective screening studies due to a lack of follow up of people who screened negative.

Positive predictive values (PPV) from four prospective screening studies using SUAC as primary marker was 100% in three studies (6 true positive cases out of 717,501 people screened),[16; 27; 32] and 67% in one study (2 true positive cases and 1 false positive case out of ~500,000 people screened).[20] PPV could not be calculated in two studies.[30; 31] There were very wide confidence intervals due to the small number of cases (in total 8 true-positive cases, ranging from 0 to three per study).

Discussion: Question 2

Study evidence

Ten studies were identified that reported test performance data for a TMS-based screening test using SUAC from DBS; six prospective screening studies and four case-control studies. The TMS methods employed differed significantly between the studies. SUAC cut-offs ranged from 1.29 µmol/l[31] to 10 µmol/l.[27] Heterogeneity in study design, TMS method used and SUAC cut-off values used, limit data synthesis. Laboratories base their SUAC cut-offs on the recovery performance of their SUAC assay. The Centers for Disease Control and Prevention's (CDC) Newborn Screening Quality Assurance Program (NSQAP) provides a proficiency testing (PT) service for laboratories that use SUAC as a marker for TYR1. Laboratories are provided with DBS specimens that are enriched with predetermined SUAC concentrations. They are instructed to provide the cut-offs they use to determine presumptive positive and negative test results, to measure the SUAC concentration of the provided specimens, and to give their clinical classifications of the specimens. [37; 38] There are large differences in the levels of SUAC identified in the enriched DBS samples among the participating screening laboratories. For example, laboratories reported SUAC concentrations of 0 to 124.90 µmol/l for DBS specimens enriched with 50 µmol/l of SUAC. Results were dependent upon the method used (nonkit TMS, kit TMS or non-TMS), the strategy used to extract DBS (freshly punched, residual DBS or co-extracted), and the internal standard and calibrators used in laboratories.[37] These method-related differences in measured SUAC concentrations of PT specimens were also observed in samples of TYR1 patients and reflect analytic biases which might explain the wide variation in the cut-off values used.

Risk of bias

Risk of bias was considered high in two or more domains in four of 10 studies (40%) and in one domain in the remaining six studies (60%). All six of the included *prospective screening studies* conducted confirmatory diagnostic tests on screen-positive samples only. Definition of clinical follow-up of screen-negative cases or losses to follow-up were not reported in any study. The time from study endpoint to manuscript submission was usually less than 14 months (too short to assess the absence of TYR1, especially the chronic form). A minimum follow-up period of at least 2 years is thought to be required to assess the presence or absence of TYR1[14] and number lost to follow-up should be reported. Therefore, a confirmed TYR1 status is only available for a minority of the population, and the true disease status of those who screened negative is not known. The determination of sensitivity, specificity, and negative predictive value is therefore not possible from these studies.

The four included *case-control studies* found a clear discrimination between affected and healthy newborns and claimed that TMS screening using SUAC as marker could achieve 100% sensitivity and specificity. Risk of bias was high for the index test domain in 3 of 4 studies (75%) as the SUAC cut-off used was study-derived and not pre-specified from an independent sample set. Unclear or high risk of verification bias was present in all case-control studies as the reference standard used to confirm presence or absence of TYR1 was not described, was inappropriate i.e. 2nd-tier screening approach by Magera et al. (2006),[39] or was not the same for cases and controls.

Applicability

The screening population itself (i.e. general newborn population) and the type of specimen used (i.e. DBS) were directly applicable to a UK screening programme in all six prospective screening studies, but three studies collected DBS samples earlier than is standard in the UK (at 1.5 to 3 days[16; 27; 30] compared to 5 to 8 days in the UK). One study[32] received samples from babies over one month of age for screening, and two studies did not report age at sampling.[20; 31] Age at sampling influences SUAC levels in babies affected by TYR1, and as a result the apparent sensitivity and specificity of the screening test. Applicability of the index test to a UK screening programme was reduced in two studies[20; 32] which used an indeterminate SUAC range and where a repeat DBS sample was requested before classification of samples as positive or negative.

The only consistent performance metric available was positive predictive value (PPV), which is not intrinsic to the test, and depends on the prevalence of TYR1 in the tested population. Data from the prospective screening studies reported a PPV of 100% in three (6 true positive cases out of 717,501 people screened)[16; 27; 32] and 67% in one study (2 true positive cases and 1 false positive case out of ~500,000 people screened)[20]. PPV could not be calculated in two studies.[30; 31]

Consistency

Results from case-control studies are consistent and promising but are not definitive and would have to be confirmed in well-designed prospective cohort studies with appropriate follow-up of screen-negative children.

Summary

Criterion 4: Not met.

Case-control studies reported clear discrimination between SUAC levels of affected and healthy newborns, but SUAC cut-offs were not pre-specified before the start of the study using an independent sample set in 3 out of 4 studies and verification bias might be present in all four studies as the reference standard used to confirm absence of TYR1 in healthy controls was not reported or was considered inappropriate (2nd-tier SUAC screening). Screening programmes using SUAC as primary marker have reported a PPV between 67% and 100% (8 true positive cases and 1 false positive case out of ~1,217,501 people screened). Sensitivity and specificity could not be determined from the included studies as duration of follow up of screen-negative cases and losses to follow-up were not reported. Differences in the test performance might be due to the timing of the test, the method used to extract SUAC, the cut-off used for classifying the disease status, dealing with flagged samples (repeat testing), or variation in normal SUAC values in the tested newborn population. TMS measurement of SUAC from dried blood spots looks like a promising screening test for TYR1 but test performance from proof-of-concept studies should be confirmed in prospective screening studies with appropriate follow-up of screen-negatives.

Table 4. Test accuracy of TMS screening for tyrosinaemia type 1 using succinylacetone as primary marker

Study	Number screened	2x2 table				Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	QUADAS-2 domain(s) with high risk of bias	QUADAS-2 domain(s) with high applicability concerns	Notes
		TP	TN	FP	FN							
Prospective screening studies using SUAC as primary marker												
La Marca 2011[30] Tuscan NBS programme Prospective NBS over ~4 years. SUAC cut-off: 2.4 µmol/l.	136,075 [Overlap of 13,000 samples with La Marca (2008)[36]]	2	NA	NA	NA	NA	NA	*NA*	NA	Flow & Timing	Patients	The authors report 'no false positive on record' but provide no details of what records checked or the extent of this checking.'
Lund 2012[16] Routine expanded NBS programme in Denmark, Faroe Islands, and Greenland. Prospective NBS over 26 months. SUAC cut-off: 2.1 U.	140,565	1	NA	0	NA	NA	NA	100 *(5.5-100)*	NA	Flow & Timing	Patients	No reported follow-up of 140,564 screen-negatives. 0 FN and 0 FP reported. False-positive rate 0% reported. All children with TYR1 are diagnosed and treated in the same centre.
Metz 2012[31] Routine Austrian NBS programme Prospective NBS over 1 month plus retrospective analysis of stored DBS from cases. SUAC cut-off: 1.29 µmol/l.	4,683 consecutively screened [plus 3 known cases representing a case-control analysis]	0 [3]	NA	0	NA	NA	NA	NA *[100 (31.0-100) using case-control analysis]*	NA	Flow & Timing	none	No reported follow-up of 4,683 screen-negatives.
Morrissey 2011[20] New York State NBS programme Prospective NBS over 24 months. SUAC cut-off: 3.00 µmol/l for retest; Average (initial and retest) 3.00-5.00 µmol/l for repeat DBS specimen request; Average (initial and retest) ≥ 5.00 µmol/l for referral.	~500,000	2	NA	1	NA	NA	NA	*66.7 (12.5-98.2)*	NA	Flow & Timing	none	2 with borderline initial & retest SUAC → repeat specimen negative. Exact number of screen-negatives NR. No reported follow-up of screen-negatives.

Study	Number screened	2x2 table				Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	QUADAS-2 domain(s) with high risk of bias	QUADAS-2 domain(s) with high applicability concerns	Notes
		TP	TN	FP	FN							
Sander 2006[27] Germany Prospective NBS over 16 weeks plus retrospective analysis of stored original DBS from cases. SUAC cut-off: 10 µmol/l.	61,344 unselected newborns [plus 2 confirmed cases]	2 [4]	NA	0	NA	NA	NA	*100 (19.8-100) [100 (39.6-100)]*	NA	Flow & Timing	Patients	Follow-up of all 61,344 newborns was not conducted
Zytkovicz 2013[32] New England NBS programme Prospective NBS over 4 years 1 month. Pooled assay: SUAC cut-off 0.55 µM → retest individually. Quantitative assay: SUAC cut-off 4 µM (recently reduced to 3.3 µM) positive; SUAC 1.0-3.3 µM indeterminate → repeat DBS request.	518,687: 515,592 newborn + 3095 older than 1 month 491,472 (94.8%) born nationally; 27,215 (5.2%) born OOC.	*3*	NA	0	NA	NA	NA	*100 (31.0-100)*	NA	Flow & Timing	Patients	5 excluded from analysis: 2 screen-positives born OOC without clinical follow-up; 3 with indeterminate SUAC born OOC without repeat DBS sample. 1 with indeterminate SUAC: repeat DBS negative. No reported follow-up of 518,679 screen-negatives. 27,209 screen-negatives born OOC difficult to follow-up. No FN known to date.

Study	Number screened	2x2 table				Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	QUADAS-2 domain(s) with high risk of bias	QUADAS-2 domain(s) with high applicability concerns	Notes
		TP	TN	FP	FN							
Case-control studies using SUAC as primary marker												
Allard 2004[33] New England NBS programme Case-control study: Residual DBS from prospective NBS for controls, retrospective analysis of stored DBS for cases. SUAC cut-off: 2 µmol/l.	4,002: 3,199 unaffected newborns and 3 known cases.	3	NA	0	NA	*100 (31.0-100)*	*100 (99.85-100)*	*100 (31.0-100)*	*100 (99.85 -100)*	Patient selection, Index test	Patients	Cut-off study-derived retrosepctively. Reference standard for healthy controls NR.
Dhillon 2011[35] California NBS programme Case-control study: Prospective NBS over 1 month for controls; retrospective analysis of stored DBS for cases. SUAC cut-off: 3 µmol/l.	~1,020: >1,000 normal babies and 6 confirmed cases.	6	>1,000	0	0	*100 (51.7-100)*	NA	*100 (51.7-100)*	NA	Patient selection, Index test	none	Cut-off study-derived retrospectively; Reference standard for healthy controls NR. Exact number of TN unclear.
La Marca 2008[36] Tuscan NBS programme Case-control study: Prospective NBS over 4 months for controls; retrospective analysis of stored DBS for cases. SUAC cut-off: 2.4 µmol/l.	13,006: 13,000 healthy controls and 10 samples from 6 confirmed cases. [Overlap with samples reported by La Marca et al. (2011)[30]]	6	13,000	0	0	*100 (51.7-100)*	*100 (99.96-100)*	*100 (51.7-100)*	*100 (99.96 -100)*	Patient selection, Flow & Timing	Patients	Reference standard for healthy controls NR.
*NBS samples only**	*13,005*	*5*	13,000	0	0	*100 (46.3-100)*	*100 (99.96-100)*	*100 (46.3-100)*	*100 (99.96 -100)*			* 5 of 10 stored DBS from NBS.

Study	Number screened	2x2 table				Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	QUADAS-2 domain(s) with high risk of bias	QUADAS-2 domain(s) with high applicability concerns	Notes
		TP	TN	FP	FN							
Turgeon 2008[34] Minnesota/USA Mayo Clinic's supplemental NBS programme. Case-control study: Stored random NBS samples not suggestive of TYR1 and stored original DBS from confirmed cases. SUAC cut-off: 5.0 µmol/l.	13,532: 13,521 controls and 11 known cases.	11	13,521	0	0	*100 (67.9-100)*	*100 (99.96-100)*	*100 (67.9-100)*	*100 (99.96-100)*	Patient selection, Index test, Reference standard, Flow & Timing	None	Cut-off study-derived retrospectively; Controls not suggestive of TYR1 using 2nd-tier SUAC screening approach by Magera et al. 2006.[39]

CI, confidence interval; DBS, dried blood spot; FN, false negative; FP, false positive; NA, not applicable; IEM, inborn errors of metabolism; NBS, newborn blood spot screening; NR, not reported; OOC, out of country; SUAC, succinylacetone; TMS, tandem mass spectrometry; TN, true negative; TP, true positive.

Numbers in italics were calculated by reviewers.

4.3. Key question 3 (Early vs. late treatment)

Does early treatment with Nitisinone (NTBC, Orfadin®) following screening provide better long-term outcomes than later treatment with Nitisinone after the presentation of symptoms?

This relates to NSC criterion 9:

'There should be an effective intervention for patients identified through screening, with evidence that intervention at a pre-symptomatic phase leads to better outcomes for the screened individual compared with usual care. Evidence relating to wider benefits of screening, for example those relating to family members, should be taken into account where available. However, where there is no prospect of benefit for the individual screened then the screening programme shouldn't be further considered.'

Description of the evidence

Figure 6 provides the PRISMA flow diagram for the Nitisinone treatment review. Our searches identified 430 unique records of which 28 full text articles were assessed. Of these, 22 articles were subsequently excluded using the pre-defined inclusion / exclusion criteria (see Appendix 4 for excluded studies with reason). This left 6 articles (reporting data from three studies) which met the inclusion criteria and which were included in the narrative synthesis.

Characteristics of included studies

Nitisinone treatment outcomes from two prospective cohorts (Birmingham study and Québec study) and one multicentre survey (see Table 6) reported in 6 papers. Descriptions of individual papers can be found in Appendix 12. One part of the Québec study reported prospective and retrospective data collection[40] while the other part of the Québec study[41], the Birmingham study[42; 43; 44] and the multicentre survey[45] reported retrospective only. The number of TYR1 patients included per paper ranged from 17 from a single centre in the UK[43] to 168 from 21 centres in Europe, Turkey and Israel.[45] One study used cross-sectional data from a survey,[45] follow-up time was up to 25 years in the Québec study,[40; 41] between one and 10 years in one part of the Birmingham study[44] and not reported in two other parts of the Birmingham study.[42; 43]

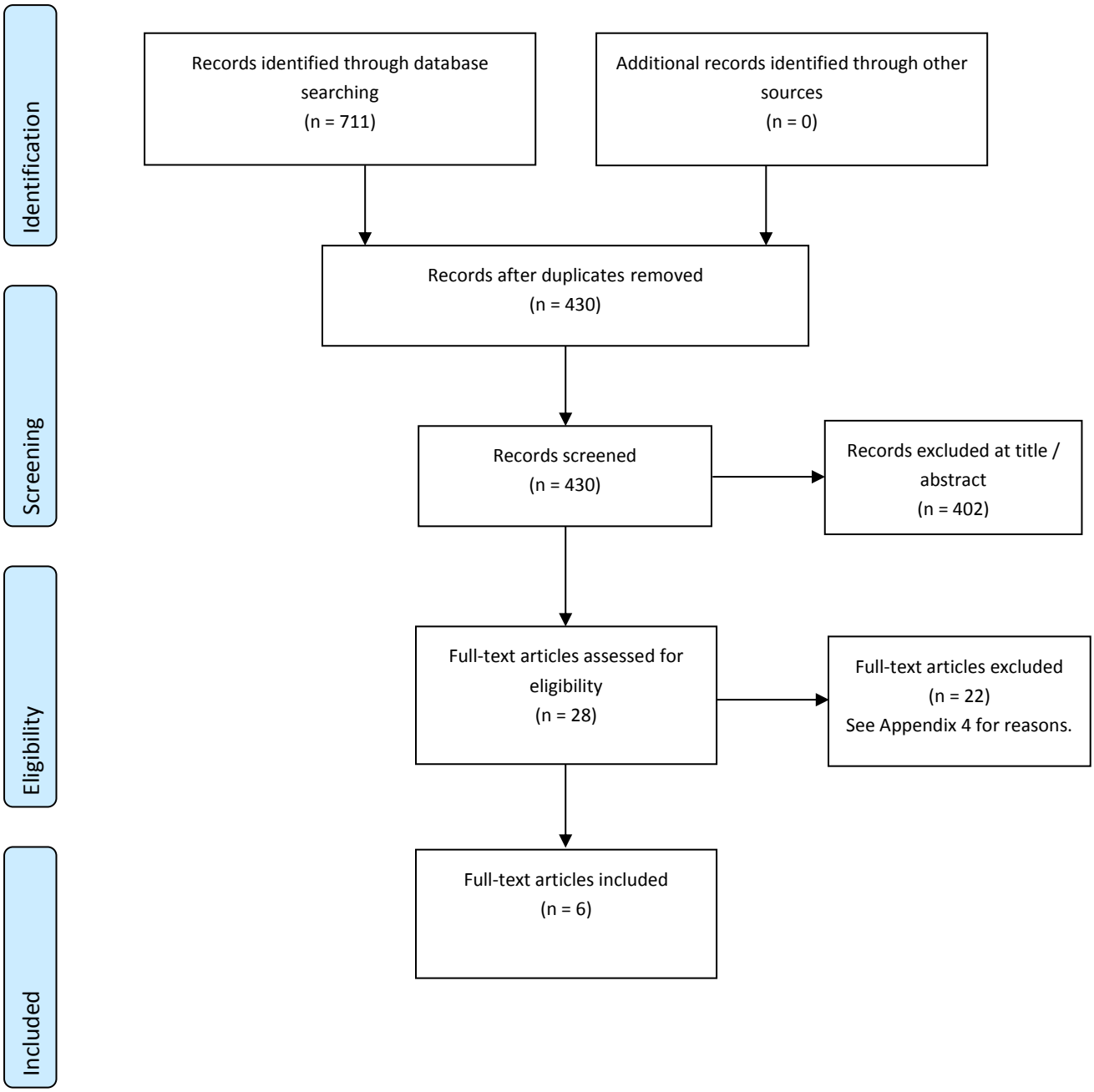

Figure 6. PRISMA Flow Diagram: Early vs. late Nitisinone treatment

Methodological quality of included studies

The methodological quality was moderate (one weak rating) in the Québec study[40; 41], moderate to weak (two or more weak ratings) in the Birmingham study[42; 43; 44] and weak in the multicentre survey[45] (see Table 5). There was high risk of selection bias in the survey[45] and in one part of the Birmingham study,[43] and all three studies had a high potential of confounding as important factors (i.e. pre-existing health problems, presenting form of TYR1, compliance with treatment, co-treatment) were not controlled in study design or analysis.

Table 5. Study quality of included studies according to EPHPP quality assessment tool[5]

Study	Global rating from sections A-F						Global rating for this study
	A) Selection bias	B) Study design	C) Confounders	D) Blinding	E) Data collection methods	F) Withdrawals and drop-outs	
Québec study							
Larochelle 2012[40]	Strong	Moderate	Weak	Moderate	Strong	Strong	Moderate
Simoncelli 2015[41]	Strong	Moderate	Weak	Moderate	Strong	Strong	Moderate
Birmingham study							
Bartlett 2014[42]	Strong	Moderate	Weak	Moderate	Strong	Strong	Moderate
McKiernan 2015[43]	Weak	Moderate	Weak	Moderate	Weak	Strong	Weak
Santra 2008[44]	Moderate	Moderate	Weak	Moderate	Weak	Weak	Weak
Survey							
Mayorandan 2014[45]	Weak	Moderate	Weak	Moderate	Weak	Strong	Weak

Analysis of the evidence

Evidence from the two cohorts and one survey (Table 6, Appendix 12) suggest that Nitisinone is an effective treatment. There is some evidence that early treatment with Nitisinone and diet may be associated with a reduction in mortality rate,[40; 41; 43] need for liver transplantation,[40; 41; 42; 45] liver cirrhosis,[45] hepatomegaly,[45] chronic liver disease,[43] rickets[45] and renal dysfunction[45]. However, this is subject to considerable bias and applicability concerns.

Discussion: Question 3

Study evidence

Overall, three studies evaluated the potential benefits of early versus late Nitisinone treatment (two cohorts and one survey). Key weaknesses across all included studies were the study design and study size: one part of the Québec study, the Birmingham study as well as the multicentre survey reported a retrospective cohort design while the other part of the Québec study reported prospective data collection for NTBC treated patients only.[40] The number of TYR1 patients included in the analysis ranged from 17 to 168. There was a high overlap in TYR1 cases between the two papers from the Québec study[40; 41] with 78 of 95 patients (82%) included in both. TYR1 cases included in the three papers of the Birmingham (UK)-based study[42; 43; 44] overlap widely (all 21 patients from Santra et al. and at least 15 of 17 patients [88%] from McKiernan et al. seem to be also included in the paper by Bartlett et al.). The TYR1 cases from Birmingham are possibly also included in the multicentre survey by Mayorandan et al. (2014)[45] that included 168 patients. Sample size per treatment group was very small in two parts of the Birmingham study[43; 44] (between five and 12 children per group) and no statistical analysis was performed.

Risk of bias

The methodological quality was moderate to weak in all three studies. There was high risk of selection bias in the multicentre survey[45] and in one part of the Birmingham study.[43] Dose of Nitisinone and dietary treatment varied between the studies and compliance with treatment was not reported. Confounding factors were not controlled for in any study, therefore benefits of early Nitisinone treatment might be due to differences in spectrum of disease, or pre-existing health problems between screen-detected/early-presenting cases and cases presenting later. Grouping unscreened patients according to age at NTBC initiation[45] or age at presentation[42] might result in differences in spectrum of disease between the groups as initial symptoms are age-dependent[45] and the clinical course differs according to age at onset of symptoms and form of TYR1.[3]

Applicability

None of the three included studies provided a comparison of outcomes of Nitisinone treatment following TYR1 detection by universal newborn screening versus treatment for an unscreened population following symptomatic presentation in the same population. In the Québec study,[40; 41] all of the patients with TYR1 in the early treatment group were identified by universal newborn screening but patients with TYR1 in the late treatment group (after 1 month of age, n=26) included 21 children (81%) who were also screen-detected but diagnosed before Nitisinone became available

and treated with diet and supportive treatment only for up to 7 years. Only five patients in this group presented clinically with symptoms and were treated with Nitisinone as soon as diagnosis was confirmed, so they are likely to have been treated significantly later than symptomatically presenting children in the UK. Larochelle et al. 2012[40] found that no patient developed neurological crisis and that no hospitalisations for acute complications of TYR1 occurred after treatment with Nitisinone was started (even if started after one month of age). The applicability of the late-treated group to the actual situation in the UK, where Nitisinone treatment is started once the diagnosis is confirmed, is low and the benefits of early treatment might have been overestimated for all outcomes. In the UK-based cohort, all[43; 44] and ten of eleven patients with TYR1 (91%)[42], respectively, in the early-treated group were identified by cascade testing or routine PKU screening and compared to patients with TYR1 who presented clinically. In the survey by Mayorandan et al.[45] the methods of diagnosis were not reported by treatment group; 28 of 168 TYR1 patients were diagnosed after newborn screening and 3 cases with prenatal diagnosis were probably among the 37 patients included in the early-treated (<1 month) group.

Consistency

Results of the two cohorts and one survey suggest that Nitisinone is an effective treatment. Potential benefits of early over late treatment have not been consistently reported.

Summary

Criterion 9: Not met.

As stated in the previous review by Bazian (2014),[2] Nitisinone is an effective treatment for TYR1. Treatment with Nitisinone in combination with a tyrosine and phenylalanine restricted diet can improve survival, liver function, renal tubular dysfunction and rickets and reduce the risk or delay the development of hepatocellular carcinoma and the need for liver transplantation. Findings from the cohort studies identified suggest that early treatment started within the first two months of life may be beneficial compared to later treatment, i.e. by reducing mortality rate and the need for liver transplantation. However, study sizes were very small with overlap of patients between reported parts of the studies. The spectrum of disease in those treated early vs late in these studies may differ and this introduces bias, and other confounding factors (i.e. pre-existing health problems) were not taken into account. The applicability of findings to the UK question of whether treatment following screen detection [or presymptomatic detection in the neonatal period] is better than treatment following symptomatic presentation or cascade testing of family members is limited. No study provided a comparison of outcomes of Nitisinone treatment initiated

following TYR1 detection by universal newborn screening versus treatment initiated following symptomatically presenting TYR1 cases in the same population.

Table 6. Association between age at treatment and various outcomes in TYR1 patients

Study	Study design	Participants	Treatment	Main findings				Global quality rating (EPHPP)[5] / Weak sections/ Applicability
				Liver disease / OLT	Death	TYR1-related hospitalisation	Other	
Québec study								
Larochelle 2012[40]	Cohort study (retrospective and prospective): Patients born 1984-2004, follow-up until 2009, death or OLT (up to 25 years). Number of centres: NR (Québec)	N=78: NTBC introduced ≤30 days: n=24 (all identified by universal NBS, 2,593 patient months). >30 days: n=26 (21 identified by universal NBS before NTBC availability; 5 not screen-detected; 535 patient months pre-NTBC, 3,138 patient months with NTBC). No NTBC: n=28 (777 patient months)	No NTBC: Diet Early- and Late-NTBC: NTBC and diet	<u>OLT</u> ≤30 days: 0/24*# >30 days: 7/26 (27%)* No NTBC: 20/28 (71%) (*p<0.001 vs No NTBC; #p<0.001 vs >30 days)	≤30 days: 0/24*# >30 days: 2/26 (8%)* (2 after OLT unrelated to TYR1) No NTBC: 10/28 (36%) (8 before, 2 after OLT) (*p<0.01 vs No NTBC; #p<0.05 vs >30 days)	<u>Neurological crisis included</u> (month with event per total patient months) ≤30 days: 0/2,593 >30 days: *43/3,673 (1.2%)* (No hospitalisations occurred during 3,138 months with NTBC treatment) No NTBC: *141/777 (18.1%)* <u>Neurological crisis excluded</u> (month with event per total patient months) ≤30 days: 0/2,593 >30 days: *17/3,673 (0.5%)* (No hospitalisations occurred during 3,138 months with NTBC treatment) No NTBC: *71/777 (9.1%)*	NR	Moderate / Confounders / All early-treated patients screen-detected. 21/26 late-treated patients screen-detected but born prior NTBC availability, 5/26 late-treated patients missed by universal NBS or born outside Québec.
Simoncelli 2015[41]	Retrospective cohort study: Patients treated between 1984-2009, follow-up until 2009 or death (up to 25 years). Number of centres: 5 (Québec)	N=95: <4 weeks: n=41 First NTBC dose: Median 13 days (IQR 11-16 days) ≥4 weeks: n=26 First NTBC dose: Median 1.0 yrs (IQR 0.4-2.2 yrs) No NTBC: n=28; [supersedes Larochelle et al. (2012)[40]]	No NTBC: Diet and "curative" OLT. Early- and Late-NTBC: NTBC and diet.	<u>OLT</u> <4 weeks: 0/41 ≥4 weeks: 7/26 (27%) No NTBC: 20/28 (71%) (p<0.001)	<4 weeks: 0/41 ≥4 weeks: 2/26 (8%) No NTBC: 10/28 (36%) (p<0.001)	<u>Hospital admissions, including PICU</u> (events per person year) <4 weeks: 0.16 ≥4 weeks: 0.41; No NTBC: 0.83 (n=21) (p<0.001). <u>Days in hospital</u> (events per person year) <4 weeks: 0.4 ≥4 weeks: 3.2 No NTBC: 7.6 (n=21) (p<0.001).	<u>Patients with neurologic crisis</u> <4 weeks: 0/41 ≥4 weeks: 5/26 (19%) (All neurologic crisis occurred before NTBC initiation.) No NTBC: 14/28 (50%) (p<0.001)	Moderate / Confounders / All early-treated patients screen-detected. 21/26 late-treated patients screen-detected but born prior NTBC availability, 5/26 late-treated patients missed by universal NBS or born outside Québec.[40]

Study	Study design	Participants	Treatment	Main findings				Global quality rating (EPHPP)[5] / Weak sections/ Applicability
				Liver disease / OLT	Death	TYR1-related hospitalisation	Other	
Multicentre survey								
Mayorandan 2014[45]	Retrospective international cohort (cross-sectional data) 21 centres in Europe, Turkey and Israel.	N=168 included in study. Way of diagnosis: 3 prenatal diagnosis, 28 NBS (12 Tyrosine, 4 SUAC, 4 Tyrosine + SUAC, 8 unknown), 132 selective screening after symptoms, 5 no data. N=148 included in analysis: NTBC start: <1 month: n=37 1-6 months: n=45 7-12 months: n=20 >12 months: n=46 Overlap Birmingham study (Bartlett et al.,[42] McKiernan et al.[43] and Santra et al.[44])	NTBC and diet	<u>OLT</u> <1 month: 3% >12 months: 26%* OR 12.7 (1.5-103)† <u>Acute liver disease</u> <1 month: 0 7-12 months: 15%* <u>Liver cancer</u> <1 month: 3% > 12 months: 26%* OR 12.7 (1.5-103)† <u>Liver cirrhosis</u> <1 month: 0 7-12 months: 35%* OR 41.6 (2.2-779.9)† >12 months: 35%* OR 40.5 (2.3-704.1)† <u>Hepatomegaly</u> <1 month: 11% 1-6 months: 29%* OR 3.3 (0.9-11.3)† 7-12 months: 35%* OR 4.4 (1.1-17.7)† >12 months: 33%* OR 3.9 (1.1-13.3)† * p<0.05 vs <1 month; † <1 month: OR=1, 95% CI in brackets.	NR	NR	<u>Rickets</u> <1 month: 0 >12 months: 20%* OR 19 (1.1-338.3)† <u>Renal dysfunction</u> <1 month: 7% >12 months: 24%* OR 5.5 (1.1-26.6)† * p<0.05 vs <1 month; † <1 month: OR=1 95% CI in brackets. All other reported outcomes were not significantly different compared to early NTBC (<1 month).	Weak / Selection bias, Confounders, Data collection methods / Not all early-treated patients screen-detected.

Study	Study design	Participants	Treatment	Main findings				Global quality rating (EPHPP)[5] / Weak sections/ Applicability
				Liver disease / OLT	Death	TYR1-related hospitalisation	Other	
Birmingham study								
Bartlett 2014[42]	Retrospective cohort study: Patients treated between 1989-2009, follow-up NR. Number of centres: 1 (Birmingham Children's Hospital, UK)	N=38: Pre-NTBC: n=7 Post-NTBC: n=31 Age at presentation: <2 months: n=11 (6 cascade testing, 4 routine PKU screening, 1 NR) 2-6 months: n=11 >6 months: n=9 Overlap with papers by McKiernan et al.[43] and Santra et al.[44]	Pre-NTBC: Diet Post-NTBC: NTBC and diet.	<u>OLT</u> <2 months: 0/11 2-6 months: 3/11 (27%) >6 months: 4/9 (44%) Pre-NTBC: 6/7 (86%) No OLT: Median age at NTBC start 52 (range 2-990) days; OLT: Median 428 (range 86-821) days. (p=0.004)	NR	NR	NR	Moderate / Confounders / 10/11 early-treated cases detected through cascade testing or routine PKU screening, 1 NR. Late-treated cases presented with symptoms.
McKiernan 2015[43]	Retrospective sibling-controlled cohort Number of centres: 1 (Birmingham Children's Hospital, UK) Follow-up NR (Age at last follow-up 6.5 weeks to 19 years)	N=17: Pre-clinically diagnosed: n=12 (7 cascade testing, 4 routine PKU screening, 1 born in a country with universal NBS); NTBC start: median 4 (range 2-52) days. Clinically diagnosed siblings: n=5 Age at presentation: Median 4 (range 1.5-17) months. Overlap with papers by Bartlett et al.[42] and Santra et al.[44]	NTBC and diet (1 child presenting clinically born before NTBC was available)	<u>OLT</u> Pre-clinically: 0/12 Clinically: 1/5 (20%) <u>Liver disease:</u> Pre-clinically: 0/12 Clinically: 2/3 (67%) surviving patients.	Pre-clinically: 0/12 Clinically: 2/5 (40%) (1 born prior NTBC availability, death before OLT; 1 born at 25 weeks gestation, died from respiratory complications of prematurity)	NR	<u>Learning difficulties:</u> Pre-clinically: 4/9 (44%) Clinically: 3/3 (100%) (2 'extra educational support', 1 'learning difficulties')	Weak / Selection bias, Confounders, Data collection methods NR. / Early-treated cases detected through cascade testing or routine PKU screening. Late-treated cases presented with symptoms.

Study	Study design	Participants	Treatment	Main findings				Global quality rating (EPHPP)[5] / Weak sections/ Applicability
				Liver disease / OLT	Death	TYR1-related hospitalisation	Other	
Santra 2008[44]	Retrospective cohort study Number of centres: 1 (Birmingham Children's Hospital, UK) Follow-up: 1-10 years.	N=21: Phenotype of liver disease at presentation: Acute liver failure: n=9 Age at presentation: Median 17 weeks (range 1 month to 2 yrs). Chronic liver disease: n=7 Age at presentation: Median 60 weeks (range 2 months to 9 yrs). Pre-clinically: n=5 (cascade testing or routine PKU screening) Age at presentation: Median <1 (range <1 to 2) weeks. Overlap with papers by McKiernan et al.[43] and Bartlett et al.[42]	NTBC and diet	NR	NR	NR	<u>Proteinuria</u> High values in all 3 groups at presentation. <u>Hypophosphataemia</u> More likely in acute than pre-clinically cases at presentation ($p<0.01$). <u>Phosphaturia</u> More excessive in acute than pre-clinically cases at presentation ($p=0.05$). <u>Phosphate supplementation</u> Pre-clinically: 0/5 Clinically: 4/16 (25%) <u>Fat-soluble vitamin supplementation</u> Pre-clinically: 3/5 (60%) Clinically: 16/16 (100%) <u>Tubular dysfunction</u> All 3 markers normalised within 1 year of NTBC and remained normal at follow-up of up to 10 yrs.	Weak / Confounders, Data collection methods NR, Withdrawals and drop-outs / Early-treated cases detected through cascade testing or routine PKU screening. Late-treated cases presented with symptoms.

CI, confidence interval; EPHPP, Effective Public Health Practice Project; IQR, interquartile range; NBS, Newborn blood spot screening; NR, not reported; NTBC, Nitisinone, Orfadin®; OLT, orthotopic liver transplantation; OR, Odds ratio; PICU, Paediatric intensive care unit; PKU, phenylketonuria; TYR1, tyrosinaemia type 1; yrs, years.

5. Overall discussion

In this report we examined three key questions relating to the effectiveness and appropriateness of newborn screening using TMS for TYR1.

1. What is the incidence of TYR1 in the UK? (Criterion 1)
2. What is the test accuracy (sensitivity, specificity, and predictive values applicable to UK prevalence) of succinylacetone measurement in dried blood spots (DBS) using tandem mass spectrometry for TYR1 screening? (Criterion 4)
3. Does early treatment with Nitisinone (NTBC, Orfadin®) following screening provide better long-term outcomes than later treatment with Nitisinone after the presentation of symptoms? (Criterion 9)

We used a rapid evidence approach and separate literature searches.

For Question 1 we found no new evidence of the incidence of TYR1 for the UK or for Western-European countries. The seven included studies reported incidence estimates for Tunisia (2 studies), United Arab Emirates, Bahrain, Lebanon, Singapore, and the United States but constraints of the applied, narrow search strategy prevent a reliable answer to this question. We considered that the NSC criterion 1 was not met.

For Question 2 we found 10 studies. The extraction methods and SUAC cut-offs differed significantly between the studies and no study was judged as low or unclear risk of bias in all four domains of assessment. We had significant concerns regarding applicability of the research to the UK population for over half of the studies. In addition we could not determine sensitivity, specificity, and negative predictive value from the prospective screening studies. The positive predictive value (PPV) from four prospective SUAC studies was 100% in three studies (6 true positive cases out of 717,501 people screened), and 67% in one study (2 true positive cases and 1 false positive case out of ~500,000 people screened). PPV could not be calculated in two studies.[30; 31] As none of the studies were conducted in the UK, the applicability of the PPV to the birth prevalence of TYR1 in the UK is reduced. We considered NSC criterion 4 not met.

For Question 3 three studies (two cohorts and one survey) reported on outcomes for TYR1 patients which varied depending on age at start of NTBC treatment. One part of the Québec study reported prospective and retrospective data collection, while all others were retrospective. The number of

included TYR1 patients per paper ranged from 17 from a single centre in the UK, to 168 from 21 centres in Europe, Turkey and Israel.

Methodological quality of the three included studies was weak to moderate and all three studies had a high potential for confounding as important factors (i.e. pre-existing health problems, presenting form of TYR1) were not taken into account. There was an overlap of participants and outcomes reported in the two parts of the Québec study, the three papers from the Birmingham study, and possibly the multicentre survey. We consider that the NSC criterion 9 was not met.

Strengths and Limitations

We built on a recent review of the relevant literature and used a systematic approach to the design of our search strategies and to inclusion and exclusion and quality assessment. We were unable to synthesise our findings numerically due to incomplete 2x2 tables for screening test studies, and heterogeneity in study design for both screening test and treatment studies. We used a rapid evidence assessment approach (REA). The UK NSC requirements for the literature search process of evidence summaries recommend a systematic approach, a minimum of three databases to be searched, and to use methods to limit the number of references retrieved which are acceptable to the review in question.[46] Because of our adoption of the REA approach, search terms were narrow, searches were limited to five databases, date limits were applied for the incidence search, and only articles in the English language were included; therefore it is possible that relevant articles might have been missed by this strategy.[2]These studies were missed by our rapid review search strategy as they did not refer to tyrosinaemia in title, abstract or key words, these only contained broader terms for the condition such as 'inborn errors of metabolism' or 'inborn errors of amino acid….'. We are also aware of at least one screening study in the German language that was not included in our review because of the language restriction. Sifting and data extraction were performed by one reviewer with a random 20% checked by a second reviewer. Therefore, there is a risk of error occurring in excluding studies and in extracting the data.

We did not perform quality appraisal for key question 1 (TYR1 incidence), and the risk of bias in these studies is therefore unknown. For the other two key questions (screening and treatment), one reviewer performed quality assessment of all studies and a second reviewer checked the findings in a random 20%. Again, this may have resulted in a risk of errors. The QUADAS-2 tool[4] was not

[2] A broadened literature search (i.e. search for 'IEM' not only 'TYR1') for key question 2 (screening test) did not find any additional screening studies but identified three more studies reporting incidence estimates of TYR1 (see Appendix 13).

tailored to our key question 2 resulting in high proportion of "unclear" ratings (see Appendix 11 and Appendix 13**Error! Reference source not found.** for the adjusted QUADAS-2 results and differences). For key question 3, we used the EPHPP[5] quality assessment tool which is not specific to a certain study design. Therefore, elements of internal and external validity may not have been adequately assessed in this tool.

This review did not investigate the value of current neonatal screening programmes for PKU and cascade testing for TYR1 detection and the proportion of TYR1 cases not detected in the neonatal period. Hutchesson et al. (1996) found that when amino acid chromatography was used as the initial screening test for PKU in Birmingham (UK) between January 1985 and March 1994, the false positive detection rate for disorders of tyrosine metabolism was 0.32% at 6 days, but had fallen to 0% (specificity 100%) by 6 weeks.[14] The sensitivity was 71% (95% confidence interval 38-100%) for detection of TYR1 (5 of 7 TYR1 cases detected). A recent study by Bartlett et al. (2014) provides information about age and clinical features at diagnosis for all people with TYR1 treated at Birmingham Children's Hospital (UK) between 1989 and 2009.[42] Ten of 31 TYR1 patients (32%) were diagnosed pre-clinically following cascade testing or routine PKU screening between 1992 and 2009; for the last 5 years of this study (2004-2009,) 8 of 13 patients (62%) were diagnosed pre-clinically.

6. Conclusions and implications for policy and practice

More research is needed to evaluate the incidence of TYR1 in the UK, as well as to examine the value of current neonatal screening programmes for PKU and cascade testing for detecting TYR1.

A research project using tandem mass spectrometry measurement of SUAC from dried blood spots with follow-up of screen-negatives for at least two years would considerably strengthen the test performance data, this could be achieved through follow-up of one of the existing cohorts described in this review.

For the treatment further investigation is needed regarding whether the TYR1 cases detected by screening represent the same spectrum of disease as those detected symptomatically and whether it is certain that all screen detected babies would become symptomatic in the absence of screening. Evidence is needed on whether improved outcomes with early administration of Nitisinone are due to the effectiveness of the drug, differences in the spectrum of disease or other confounding factors. Furthermore, it is not clear whether these improved outcomes with early detection are applicable to the UK question (i.e. whether the early detected cases are sufficiently similar to SUAC screen

detected in a potential UK programme and whether the late detected cases are sufficiently similar to symptomatically detected tyrosinemia in the UK).

References

1. de Laet C, Dionisi-Vici C, Leonard JV *et al.* (2013) Recommendations for the management of tyrosinaemia type 1. *Orphanet journal of rare diseases* **8**, 8.

2. Bazian L (2014) *Screening for Tyrosinaemia I. External review against programme appraisal criteria for the UK National Screening Committee (UK NSC)*. legacy.screening.nhs.uk/policydb_download.php?doc=474: NSC UK National Screening Committee.

3. van Spronsen FJ, Thomasse Y, Smit GP *et al.* (1994) Hereditary tyrosinemia type I: a new clinical classification with difference in prognosis on dietary treatment. *Hepatology* **20**, 1187-1191.

4. Whiting PF, Rutjes AW, Westwood ME *et al.* (2011) QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of internal medicine* **155**, 529-536.

5. Thomas BH, Ciliska D, Dobbins M *et al.* (2004) A process for systematically reviewing the literature: providing the research evidence for public health nursing interventions. *Worldviews on evidence-based nursing / Sigma Theta Tau International, Honor Society of Nursing* **1**, 176-184.

6. Hadj-Taieb S, Nasrallah F, Hammami MB *et al.* (2012) Aminoacidopathies and organic acidurias in Tunisia: a retrospective survey over 23 years. *Tunis Med* **90**, 258-261.

7. Nasrallah F, Hammami MB, Ben Rhouma H *et al.* (2015) Clinical and Biochemical Profile of Tyrosinemia Type 1 in Tunisia. *Clinical laboratory* **61**, 487-492.

8. Al-Shamsi A, Hertecant JL, Al-Hamad S *et al.* (2014) Mutation Spectrum and Birth Prevalence of Inborn Errors of Metabolism among Emiratis: A study from Tawam Hospital Metabolic Center, United Arab Emirates. *Sultan Qaboos Univ Med J* **14**, e42-49.

9. Lim JS, Tan ES, John CM *et al.* (2014) Inborn error of metabolism (IEM) screening in Singapore by electrospray ionization-tandem mass spectrometry (ESI/MS/MS): An 8year journey from pilot to current program. *Molecular Genetics and Metabolism* **113**, 53-61.

10. Golbahar J, Al-Jishi EA, Altayab DD *et al.* (2013) Selective newborn screening of inborn errors of amino acids, organic acids and fatty acids metabolism in the Kingdom of Bahrain. *Mol Genet Metab* **110**, 98-101.

11. Khneisser I, Adib S, Assaad S *et al.* (2015) Cost-benefit analysis: Newborn screening for inborn errors of metabolism in Lebanon. *Journal of Medical Screening* **22**, 182-186.

12. Therrell BL, Jr., Lloyd-Puryear MA, Camp KM *et al.* (2014) Inborn errors of metabolism identified via newborn screening: Ten-year incidence data and costs of nutritional interventions for research agenda planning. *Molecular Genetics and Metabolism* **113**, 14-26.

13. Hutchesson AC, Bundey S, Preece MA *et al.* (1998) A comparison of disease and gene frequencies of inborn errors of metabolism among different ethnic groups in the West Midlands, UK. *Journal of medical genetics* **35**, 366-370.

14. Hutchesson AC, Hall SK, Preece MA *et al.* (1996) Screening for tyrosinaemia type I. *Archives of disease in childhood Fetal and neonatal edition* **74**, F191-194.

15. Bliksrud YT, Brodtkorb E, Backe PH *et al.* (2012) Hereditary tyrosinaemia type i in Norway: Incidence and three novel small deletions in the fumarylacetoacetase gene. *Scandinavian Journal of Clinical and Laboratory Investigation* **72**, 369-373.

16. Lund AM, Hougaard DM, Simonsen H *et al.* (2012) 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* **107**, 281-293.

17. Couce ML, Castineiras DE, Boveda MD *et al.* (2011) Evaluation and long-term follow-up of infants with inborn errors of metabolism identified in an expanded screening programme. *Mol Genet Metab* **104**, 470-475.

18. Lindner M, Gramer G, Haege G *et al.* (2011) Efficacy and outcome of expanded newborn screening for metabolic diseases--report of 10 years from South-West Germany. *Orphanet journal of rare diseases* **6**, 44.

19. Schulze A, Lindner M, Kohlmuller D *et al.* (2003) Expanded newborn screening for inborn errors of metabolism by electrospray ionization-tandem mass spectrometry: results, outcome, and implications. *Pediatrics* **111**, 1399-1406.

20. Morrissey MA, Sunny S, Fahim A *et al.* (2011) Newborn screening for Tyr-I: Two years' experience of the New York State program. *Molecular Genetics and Metabolism* **103**, 191-192.

21. Kasper DC, Ratschmann R, Metz TF *et al.* (2010) The national Austrian newborn screening program - eight years experience with mass spectrometry. past, present, and future goals. *Wiener klinische Wochenschrift* **122**, 607-613.

22. Wilcken B, Haas M, Joy P *et al.* (2009) Expanded newborn screening: outcome in screened and unscreened patients at age 6 years. *Pediatrics* **124**, e241-248.

23. la Marca G, Malvagia S, Casetta B *et al.* (2008) 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* **31 Suppl 2**, S395-404.

24. Masurel-Paulet A, Poggi-Bach J, Rolland MO *et al.* (2008) NTBC treatment in tyrosinaemia type I: Long-term outcome in French patients. *Journal of Inherited Metabolic Disease* **31**, 81-87.

25. Feuchtbaum L, Lorey F, Faulkner L *et al.* (2006) California's experience implementing a pilot newborn supplemental screening program using tandem mass spectrometry. *Pediatrics* **117**, S261-269.

26. Frazier DM, Millington DS, McCandless SE *et al.* (2006) The tandem mass spectrometry newborn screening experience in North Carolina: 1997-2005. *J Inherit Metab Dis* **29**, 76-85.

27. Sander J, Janzen N, Peter M *et al.* (2006) Newborn screening for hepatorenal tyrosinemia: Tandem mass spectrometric quantification of succinylacetone. *Clinical Chemistry* **52**, 482-487.

28. Comeau AM, Larson C, Eaton RB (2004) Integration of new genetic diseases into statewide newborn screening: New England experience. *American journal of medical genetics Part C, Seminars in medical genetics* **125C**, 35-41.

29. Zytkovicz TH, Fitzgerald EF, Marsden D *et al.* (2001) Tandem mass spectrometric analysis for amino, organic, and fatty acid disorders in newborn dried blood spots: a two-year summary from the New England Newborn Screening Program. *Clin Chem* **47**, 1945-1955.

30. la Marca G, Malvagia S, Pasquini E *et al.* (2011) Newborn Screening for Tyrosinemia Type I: Further Evidence that Succinylacetone Determination on Blood Spot Is Essential. *JIMD rep* **1**, 107-109.

31. Metz TF, Mechtler TP, Merk M *et al.* (2012) Evaluation of a novel, commercially available mass spectrometry kit for newborn screening including succinylacetone without hydrazine. *Clinica Chimica Acta* **413**, 1259-1264.

32. Zytkovicz TH, Sahai I, Rush A *et al.* (2013) Newborn screening for hepatorenal tyrosinemia-I by tandem mass spectrometry using pooled samples: A four-year summary by the New England newborn screening program. *Clinical Biochemistry* **46**, 681-684.

33. Allard P, Grenier A, Korson MS *et al.* (2004) Newborn screening for hepatorenal tyrosinemia by tandem mass spectrometry: Analysis of succinylacetone extracted from dried blood spots. *Clinical Biochemistry* **37**, 1010-1015.

34. Turgeon C, Magera MJ, Allard P *et al.* (2008) Combined newborn screening for succinylacetone, amino acids, and acylcarnitines in dried blood spots. *Clinical Chemistry* **54**, 657-664.

35. Dhillon KS, Bhandal AS, Aznar CP *et al.* (2011) Improved tandem mass spectrometry (MS/MS) derivatized method for the detection of tyrosinemia type I, amino acids and acylcarnitine disorders using a single extraction process. *Clinica Chimica Acta* **412**, 873-879.

36. La Marca G, Malvagia S, Pasquini E *et al.* (2008) The inclusion of succinylacetone as marker for tyrosinemia type I in expanded newborn screening programs. *Rapid Communications in Mass Spectrometry* **22**, 812-818.

37. Adam BW, Lim TH, Hall EM *et al.* (2009) Preliminary proficiency testing results for succinylacetone in dried blood spots for newborn screening for tyrosinemia type I. *Clinical Chemistry* **55**, 2207-2213.

38. Adam BW, Hall EM, Meredith NK *et al.* (2012) Performance of succinylacetone assays and their associated proficiency testing outcomes. *Clinical Biochemistry* **45**, 1658-1663.

39. Magera MJ, Gunawardena ND, Hahn SH *et al.* (2006) Quantitative determination of succinylacetone in dried blood spots for newborn screening of tyrosinemia type I. *Molecular Genetics and Metabolism* **88**, 16-21.

40. Larochelle J, Alvarez F, Bussieres JF *et al.* (2012) Effect of nitisinone (NTBC) treatment on the clinical course of hepatorenal tyrosinemia in Quebec. *Molecular Genetics and Metabolism* **107**, 49-54.

41. Simoncelli M, Samson J, Bussieres JF *et al.* (2015) Cost-Consequence Analysis of Nitisinone for Treatment of Tyrosinemia Type I. *Can J Hosp Pharm* **68**, 210-217.

42. Bartlett DC, Lloyd C, McKiernan PJ *et al.* (2014) Early nitisinone treatment reduces the need for liver transplantation in children with tyrosinaemia type 1 and improves post-transplant renal function. *Journal of Inherited Metabolic Disease* **37**, 745-752.

43. McKiernan PJ, Preece MA, Chakrapani A (2015) Outcome of children with hereditary tyrosinaemia following newborn screening. *Archives of Disease in Childhood* **100**, 738-741.

44. Santra S, Preece MA, Hulton SA *et al.* (2008) Renal tubular function in children with tyrosinaemia type I treated with nitisinone. *Journal of Inherited Metabolic Disease* **31**, 399-402.

45. Mayorandan S, Meyer U, Gokcay G *et al.* (2014) Cross-sectional study of 168 patients with hepatorenal tyrosinaemia and implications for clinical practice. *Orphanet journal of rare diseases* **9**.

46. Public Health England (2015) Appendix G: Literature searches for evidence summaries. https://www.gov.uk/government/publications/uk-nsc-evidence-review-process/appendix-g-literature-searches-for-evidence-summaries (accessed 14 March 2016)

Appendices

Appendix 1. Search strategy for Ovid Medline

Search strategies were developed for MEDLINE (Ovid) and were adapted appropriately for other databases: MEDLINE In-Process & Other Non-Indexed Citations (Ovid), EMBASE (Ovid), Cochrane Library and Web of Science.

A - Key question 1 (Prevalence search):

Searches	Results	Search Type	Actions
☐	1	exp Tyrosinemias/	309
☐	2	(tyrosinemia* or tyrosinaemia*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]	999
☐	3	(tyr1 or tyr-1).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]	963
☐	4	(tyri or tyr-i).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]	29
☐	5	(((fumarylacetoacetate adj hydrolase) or fumarylacetoacetase or fah) adj2 deficien*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]	127
☐	6	1 or 2 or 3 or 4 or 5	2008
☐	7	exp Epidemiologic Studies/	1823425
☐	8	epidemiolog*.mp.	341999
☐	9	exp Prevalence/	211721
☐	10	exp Incidence/	192736
☐	11	(prevalen* or inciden*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]	1136641
☐	12	7 or 8 or 9 or 10 or 11	2791405
☐	13	6 and 12	140
☐	**14**	**limit 13 to yr="2012 -Current"**	**19**

B - Key question 2 (Screening test search):

Searches	Results	Search Type	Actions
☐	1	exp Tyrosinemias/	309
☐	2	(tyrosinemia* or tyrosinaemia*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]	999
☐	3	(tyr1 or tyr-1).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]	963
☐	4	(tyri or tyr-i).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]	29
☐	5	(((fumarylacetoacetate adj hydrolase) or fumarylacetoacetase or fah) adj2 deficien*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]	127
☐	6	1 or 2 or 3 or 4 or 5	2008
☐	7	suac.mp.	24
☐	8	exp Heptanoates/ or succinylacetone.mp.	420
☐	9	succinylacetoacetate.mp.	14
☐	10	51568 18 4 succinylacetone.rn.	225
☐	11	4,6-dioxoheptanoate.mp.	4
☐	12	exp Heptanoic Acids/ or 4,6-Dioxoheptanoic acid.mp.	6020
☐	13	7 or 8 or 9 or 10 or 11 or 12	6140
☐	**14**	**6 and 13**	**165**

C - Key question 3 (Nitisinone treatment search):

Searches	Results	Search Type	Actions
☐	1	exp Tyrosinemias/	309
☐	2	(tyrosinemia* or tyrosinaemia*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]	999
☐	3	(tyr1 or tyr-1).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]	963
☐	4	(tyri or tyr-i).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]	29
☐	5	(((fumarylacetoacetate adj hydrolase) or fumarylacetoacetase or fah) adj2 deficien*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]	127
☐	6	1 or 2 or 3 or 4 or 5	2008
☐	7	Nitisinone.mp.	144
☐	8	104206 65 7 Nitisinone.rn.	127
☐	9	ntbc.mp.	120
☐	10	orfadin.mp.	4
☐	11	2 nitro 4 trifluoromethylbenzoyl.mp.	45
☐	12	7 or 8 or 9 or 10 or 11	181
☐	**13**	**6 and 12**	**140**

Appendix 2. Excluded studies (Incidence search) with reason (n=10)

Reference	Reason for exclusion
1. Al Riyami S, Al Maney M, Joshi SN, Bayoumi R. Detection of inborn errors of metabolism using tandem mass spectrometry among high-risk Omani patients. Oman Medical Journal. 2012;27(6):482-5.	High-risk population
2. Al-Thihli K, Al-Murshedi F, Al-Hashmi N, Al-Mamari W, Islam MM, Al-Yahyaee SA. Consanguinity, endogamy and inborn errors of metabolism in Oman: A cross-sectional study. Human Heredity. 2014;77(1-4):183-8.	No birth prevalence data for TYR1
3. Amelina SS, Vetrova NV, Amelina MA, Degtereva EV, Ponomareva TI, Elchinova GI, et al. The load and diversity of hereditary diseases in four raions of Rostov oblast. Russian Journal of Genetics. 2014;50(1):82-90.	No birth prevalence data for TYR1
4. Angileri F, Bergeron A, Morrow G, Lettre F, Gray G, Hutchin T, et al. Geographical and Ethnic Distribution of Mutations of the Fumarylacetoacetate Hydrolase Gene in Hereditary Tyrosinemia Type 1. JIMD rep. 2015;19:43-58.	No new birth prevalence data; Bliksrud 2012 already included in Bazian review
5. Bliksrud YT, Brodtkorb E, Backe PH, Woldseth B, Rootwelt H. Hereditary tyrosinaemia type i in Norway: Incidence and three novel small deletions in the fumarylacetoacetase gene. Scandinavian Journal of Clinical and Laboratory Investigation. 2012;72(5):369-73.	Already included in Bazian review
6. De Jesus VR, Adam BW, Mandel D, Cuthbert CD, Matern D. Succinylacetone as primary marker to detect tyrosinemia type I in newborns and its measurement by newborn screening programs. Molecular Genetics and Metabolism. 2014;113(1):67-75.	No birth prevalence data for TYR1
7. Han L, Han F, Ye J, Qiu W, Zhang H, Gao X, et al. Spectrum Analysis of Common Inherited Metabolic Diseases in Chinese Patients Screened and Diagnosed by Tandem Mass Spectrometry. Journal of Clinical Laboratory Analysis. 2015;29(2):162-8.	High-risk population
8. Mak CM, Lee HCH, Chan AYW, Lam CW. Inborn errors of metabolism and expanded newborn screening: Review and update. Critical Reviews in Clinical Laboratory Sciences. 2013;50(6):142-62.	No birth prevalence data for TYR1
9. Nakamura K, Matsumoto S, Mitsubuchi H, Endo F. Diagnosis and treatment of hereditary tyrosinemia in Japan. Pediatrics International. 2015;57(1):37-40.	No birth prevalence data for TYR1
10. Shawky RM, Abd-Elkhalek HS, Elakhdar SE. Selective screening in neonates suspected to have inborn errors of metabolism. Egyptian Journal of Medical Human Genetics. 2015;16(2):165-71.	High-risk population

Appendix 3. Excluded studies (SUAC search) with reason (n=10)

Reference		Reason for exclusion
1.	Adam BW, Hall EM, Meredith NK, Lim TH, Haynes CA, De Jesus VR, et al. Performance of succinylacetone assays and their associated proficiency testing outcomes. Clinical Biochemistry. 2012;45(18):1658-63.	No test performance data for TYR1 screening
2.	Adam BW, Lim TH, Hall EM, Hannon WH. Preliminary proficiency testing results for succinylacetone in dried blood spots for newborn screening for tyrosinemia type I. Clinical Chemistry. 2009;55(12):2207-13.	No test performance data for TYR1 screening
3.	Al-Dirbashi OY, Rashed MS, Brink HJ, Jakobs C, Filimban N, Al-Ahaidib LY, et al. Determination of succinylacetone in dried blood spots and liquid urine as a dansylhydrazone by liquid chromatography tandem mass spectrometry. J Chromatogr B Analyt Technol Biomed Life Sci. 2006;831(1-2):274-80.	HPLC-MS/MS method not suitable for universal NBS (primary screen)
4.	Al-Dirbashi OY, Rashed MS, Jacob M, Al-Ahaideb LY, Al-Amoudi M, Rahbeeni Z, et al. Improved method to determine succinylacetone in dried blood spots for diagnosis of tyrosinemia type 1 using UPLC-MS/MS. Biomedical Chromatography. 2008;22(11):1181-5.	No cut-off for and test performance data reported
5.	De Jesus VR, Adam BW, Mandel D, Cuthbert CD, Matern D. Succinylacetone as primary marker to detect tyrosinemia type I in newborns and its measurement by newborn screening programs. Molecular Genetics and Metabolism. 2014;113(1):67-75.	No test performance data for TYR1 screening
6.	Johnson DW, Gerace R, Ranieri E, Trinh MU, Fingerhut R. Analysis of succinylacetone, as a Girard T derivative, in urine and dried bloodspots by flow injection electrospray ionization tandem mass spectrometry. Rapid Communications in Mass Spectrometry. 2007;21(1):59-63.	No cut-off and test performance data, method development
7.	Magera MJ, Gunawardena ND, Hahn SH et al. (2006) Quantitative determination of succinylacetone in dried blood spots for newborn screening of tyrosinemia type I. Molecular Genetics and Metabolism 88, 16-21.	SUAC as 2nd-tier test
8.	Marca GL, Malvagia S, Funghini S, Pasquini E, Moneti G, Guerrini R, et al. The successful inclusion of succinylacetone as a marker of tyrosinemia type i in Tuscany newborn screening program. Rapid Communications in Mass Spectrometry. 2009;23(23):3891-3.	Letter, case report of a true positive case
9.	Matern D, Tortorelli S, Oglesbee D et al. (2007) Reduction of the false-positive rate in newborn screening by implementation of MS/MS-based second-tier tests: The Mayo Clinic experience (2004-2007). Journal of Inherited Metabolic Disease 30, 585-592.	SUAC as 2nd-tier test
10.	McHugh DMS, Cameron CA, Abdenur JE, Abdulrahman M, Adair O, Al Nuaimi SA, et al. Clinical validation of cut-off target ranges in newborn screening of metabolic disorders by tandem mass spectrometry: A worldwide collaborative project. Genetics in Medicine. 2011;13(3):230-54.	No test performance data for TYR1 screening

Appendix 4. Excluded studies (Nitisinone search) with reason (n=22)

Reference	Reason for exclusion
1. Alobaidy HA, Yahya NA, Said RM. Tyrosinemia type 1: Clinical and biochemical analysis of cases with poor treatment outcome. Jordan Medical Journal. 2011;45(2):205-12.	Case reports of 3 TYR1 cases
2. Anonymous. Nitisinone. Type 1 tyrosinemia: An effective drug. Prescrire International. 2007;16(88):56-8.	Duplicate
3. Anonymous. Nitisinone. Australian Prescriber. 2009;32(2):54-5.	No early vs late NTBC comparison
4. Anonymous. Nitisinone: new drug. Type 1 tyrosinemia: an effective drug. Prescrire International. 2007;16(88):56-8.	Not a systematic review
5. Arora N, Stumper O, Wright J, Kelly DA, McKiernan PJ. Cardiomyopathy in tyrosinaemia type I is common but usually benign. Journal of Inherited Metabolic Disease. 2006;29(1):54-7.	No early vs late NTBC comparison
6. Baumann U, Rodeck B. Liver transplantation in tyrosinaemia type I. Monatsschrift Kinderheilkunde. 2004;152(10):1102-6.	German language
7. Buckley BM. Clinical trials of orphan medicines. The Lancet. 2008;371(9629):2051-5.	No early vs late NTBC comparison
8. De Laet C, Terrones Munoz V, Jaeken J, Francois B, Carton D, Sokal EM, et al. Neuropsychological outcome of NTBC-treated patients with tyrosinaemia type 1. Developmental Medicine and Child Neurology. 2011;53(10):962-4.	Letter, no early vs late NTBC comparison
9. Elpeleg ON, Shaag A, Holme E, Zughayar G, Ronen S, Fisher D, et al. Mutation analysis of the FAH gene in Israeli patients with tyrosinemia type I. Human mutation. 2002;19(1):80-1.	No early vs late NTBC comparison
10. Gissen P, Preece MA, Willshaw HA, McKiernan PJ. Ophthalmic follow-up of patients with tyrosinaemia type I on NTBC. Journal of Inherited Metabolic Disease. 2003;26(1):13-6.	No early vs late NTBC comparison
11. Holme E, Lindstedt PS, Lock EA. Treatment of tyrosinemia type I with an enzyme inhibitor (NTBC). International Pediatrics. 1995;10(1):41-3.	No early vs late NTBC comparison
12. Holme E, Lindstedt S. Tyrosinaemia type I and NTBC (2-(2-nitro-4-trifluoromethylbenzoyl)-1,3- cyclohexanedione). Journal of Inherited Metabolic Disease. 1998;21(5):507-17.	No early NTBC group following screening; only NTBC before/after 2 years of age
13. Holme E, Lindstedt S. Nontransplant treatment of tyrosinemia. Clinics in Liver Disease. 2000;4(4):805-14.	Excluded as no early (screened) vs late NTBC data
14. Joshi SN, Venugopalan P. Experience with NTBC therapy in hereditary tyrosinaemia type I: An alternative to liver transplantation. Annals of Tropical Paediatrics. 2004;24(3):259-65.	No early vs late NTBC comparison
15. Kitagawa T. Hepatorenal tyrosinemia. Proceedings of the Japan Academy Series B: Physical and Biological Sciences. 2012;88(5):192-200.	No systematic review, no early vs late NTBC comparison
16. Masurel-Paulet A, Poggi-Bach J, Rolland MO, Bernard O, Guffon N, Dobbelaere D, et al. NTBC treatment in tyrosinaemia type I: Long-term outcome in French patients. Journal of Inherited Metabolic Disease. 2008;31(1):81-7.	Early NTBC group (<6 months) is not screen-detected; comparison of acute v. sub-acute vs chronic forms of TYR1
17. McKiernan PJ. Nitisinone in the treatment of hereditary tyrosinaemia type 1. Drugs. 2006;66(6):743-50.	No early vs late NTBC comparison
18. McKiernan PJ, Preece MA, Green A, Lindstedt S, Holme E, Lock EA, et al. IMPROVEMENT IN LIVER-FUNCTION AND HISTOLOGY IN TYROSINEMIA TYPE-1 WITH NTBC. Hepatology. 1995;22(4):1076-.	Conference abstract
19. Nakamura K, Matsumoto S, Mitsubuchi H, Endo F. Diagnosis and treatment of hereditary tyrosinemia in Japan. Pediatrics International. 2015;57(1):37-40.	No early vs late NTBC comparison
20. Pierre G, Chronopoulou E. Metabolic disorders presenting as liver disease. Paediatrics and Child Health (United Kingdom). 2013;23(12):509-15.	No early vs late NTBC comparison
21. Van Spronsen FJ, Bijleveld CMA, Van Maldegem BT, Wijburg FA. Hepatocellular carcinoma in hereditary tyrosinemia type I despite 2-(2 nitro-4-3 trifluoro- methylbenzoyl)-1, 3-cyclohexanedione treatment. Journal of Pediatric Gastroenterology and Nutrition. 2005;40(1):90-3.	Case report; no early vs late NTBC comparison
22. Wijburg FA, Reitsma Ch WC, Slooff MJH, Van Spronsen FJ, Koetse HA, Reijngoud DJ, et al. Liver transplantation in tyrosinaemia type I: The Groningen experience. Journal of Inherited Metabolic Disease. 1995;18(2):115-8.	No early vs late NTBC comparison

Appendix 5. Data extraction form for included studies

Data extraction form for primary studies

Name of first reviewer: *Name of second reviewer:*

Study details				
Study ID (Endnote ref)				
First author surname and year of publication				
Country				
Study design				
Study setting				
Number of centres				
Time period / study duration				
Follow up period				
Funding				
Competing interests / Role of sponsor				
Review question				
Prevalence (1) Screening test (2) NTBC treatment (3)				
Aim of the study				
Patient selection				
Inclusion criteria:				
Exclusion criteria:				
Study flow				
Item	**All**	**Early NTBC**	**Late NTBC**	**Never NTBC**
Screened				
Randomised/Included in study				
Excluded from study (reasons)				
Missing participants				
Withdrawals				
Included in analysis				
Excluded from analysis (reasons)				

Baseline characteristics				
For 1-3)	**All**	**Early NTBC**	**Late NTBC**	**Never NTBC**
Total number of participants				
Age Mean (SD) Median (range) years				
Sex (male/female, % female)				
Ethnicity				
Education: Regular Special				
Socioeconomic status: Low Middle High				
Comments on the presence or absence of significant differences between treatment arms:				
Additional baseline characteristics				
For 3) NTBC treatment	**All**	**Early NTBC**	**Late NTBC**	**Never NTBC**
Total number of participants				
Method of diagnosis				
Age at diagnosis, median (range)				
Pathogenic FAH mutation				
Age at start of NTBC, median (range)				
Patient-months without NTBC treatment				
Patient-months with NTBC treatment				
Current NTBC use (y/n)				
Consanguinity (y/n)				
Affected family members (y/n)				

Interventions & comparators			
1) Prevalence	**Screened cohort**	**Unscreened cohort**	**Historic cohort**
Screening programme in operation? (y/n) If yes, fill 2) for details			
Confirmation of disease			
Follow-up (years)			
2) TMS screening method			
Source and type of material			
Age at specimens collection			
Sample transport and storage			
Samples pooled?			
Method of extraction & TMS analysis			
Type of tandem MS			
Data management			
Quality assurance			
Analysis			
Cut-off / Threshold			
Cut-off prespecified (y/n)			
Positive screening results			
Reference standard used			
Follow-up (years)			
Number received index test, n (%)			
Number received reference standard			
3) NTBC treatment			
	Early NTBC	**Late NTBC**	**Never NTBC**
Total number			
Age at start of NTBC treatment, median (range)			
NTBC dose (mg/kg/day)			
NTBC frequency			
Other medication			
Dietary treatment			
Monitoring			
Follow-up (patient months)			
IQ test used			
Other (specify)			

Outcomes			
1) Prevalence / Incidence of TYR1			
Reported outcomes:			
	Screened cohort	**Unscreened cohort**	**Historic cohort**
Number screened			
Number of identified cases			
Incidence			
Cases per 100,000			
Notes / Comments:			

2) Screening for TYR1	
Reported outcomes:	
Total number screened	
TP	
TN	
FP	
FN	
Sensitivity, % (95% CI)	
Specificity, % (95% CI)	
PPV, % (95% CI)	
NPV, % (95% CI)	
Other (specify)	
Notes / Comments:	

3) NTBC treatment			
Reported outcomes:			
	Early NTBC	**Late NTBC**	**Never NTBC**
Number included patients			
Death			
Death before liver transplantation (LT)			
Death after LT			
LT			
Age at LT			
Months with TYR1-related hospitalisations (neurological crises included)			
Months with neurological crises			
Acute liver failure			
Chronic liver disease			
Carcinoma			
Cirrhosis			
Hepatomegaly			
Rickets			
Renal dysfunction			
Renal tubular dysfunction			

Nephromegaly			
Nephrocalcinosis			
Neurological crisis			
ADS, behavioural disorders			
Learning/language difficulties			
Impaired psychomotor development			
IQ			
Plasma tyrosine level			
Plasma phenylalanine level			
Others (specify):			
Notes / comments:			
Authors' comments & conclusion			
Reviewer's comments & conclusion			

Appendix 6. Quality assessment forms

A – QUADAS-2 tool with index questions adapted to the review for studies comparing performance of different tests[4]

QUADAS-2

Phase 1: State the review question:

Patients (setting, intended use of index test, presentation, prior testing):
Index test(s):
Reference standard and target condition:

Phase 2: Draw a flow diagram for the primary study

Phase 3: Risk of bias and applicability judgments

QUADAS-2 is structured so that 4 key domains are each rated in terms of the risk of bias and the concern regarding applicability to the research question (as defined above). Each key domain has a set of signalling questions to help reach the judgments regarding bias and applicability.

DOMAIN 1: PATIENT SELECTION

A. Risk of Bias

Describe methods of patient selection:

- Was a consecutive or random sample of patients enrolled? Yes/No/Unclear
- Was a case-control design avoided? Yes/No/Unclear
- Did the study avoid inappropriate exclusions? Yes/No/Unclear

Could the selection of patients have introduced bias? **RISK: LOW/HIGH/UNCLEAR**

B. Concerns regarding applicability

Describe included patients (prior testing, presentation, intended use of index test and setting):

Is there concern that the included patients do not match the review question? **CONCERN: LOW/HIGH/UNCLEAR**

DOMAIN 2: INDEX TEST(S)

If more than one index test was used, please complete for each test.

A. Risk of Bias

Describe the index test and how it was conducted and interpreted:

- Were the index test results interpreted without knowledge of the results of the reference standard? Yes/No/Unclear
- If a threshold was used, was it pre-specified? Yes/No/Unclear

Could the conduct or interpretation of the index test have introduced bias? **RISK: LOW /HIGH/UNCLEAR**

B. Concerns regarding applicability

Is there concern that the index test, its conduct, or interpretation differ from the review question? **CONCERN: LOW /HIGH/UNCLEAR**

DOMAIN 3: REFERENCE STANDARD

A. Risk of Bias

Describe the reference standard and how it was conducted and interpreted:

- ❖ Is the reference standard likely to correctly classify the target condition? Yes/No/Unclear
- ❖ Were the reference standard results interpreted without knowledge of the results of the index test? Yes/No/Unclear

Could the reference standard, its conduct, or its interpretation have introduced bias? **RISK: LOW /HIGH/UNCLEAR**

B. Concerns regarding applicability

Is there concern that the target condition as defined by the reference standard does not match the review question? **CONCERN: LOW /HIGH/UNCLEAR**

DOMAIN 4: FLOW AND TIMING

A. Risk of Bias

Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2x2 table (refer to flow diagram):

Describe the time interval and any interventions between index test(s) and reference standard:

- ❖ Was there an appropriate interval between index test(s) and reference standard? Yes/No/Unclear
- ❖ Did all patients receive a reference standard? Yes/No/Unclear
- ❖ Did patients receive the same reference standard? Yes/No/Unclear
- ❖ Were all patients included in the analysis? Yes/No/Unclear

Could the patient flow have introduced bias? **RISK: LOW /HIGH/UNCLEAR**

B – EPHPP quality assessment tool for quantitative studies[5]

QUALITY ASSESSMENT TOOL FOR QUANTITATIVE STUDIES

EPHPP
Effective Public Health Practice Project

COMPONENT RATINGS

A) SELECTION BIAS

(Q1) Are the individuals selected to participate in the study likely to be representative of the target population?

1 Very likely
2 Somewhat likely
3 Not likely
4 Can't tell

(Q2) What percentage of selected individuals agreed to participate?

1 80 - 100% agreement
2 60 – 79% agreement
3 less than 60% agreement
4 Not applicable
5 Can't tell

RATE THIS SECTION	STRONG	MODERATE	WEAK
See dictionary	1	2	3

B) STUDY DESIGN

Indicate the study design

1 Randomized controlled trial
2 Controlled clinical trial
3 Cohort analytic (two group pre + post)
4 Case-control
5 Cohort (one group pre + post (before and after))
6 Interrupted time series
7 Other specify _____
8 Can't tell

Was the study described as randomized? If NO, go to Component C.

No Yes

If Yes, was the method of randomization described? (See dictionary)

No Yes

If Yes, was the method appropriate? (See dictionary)

No Yes

RATE THIS SECTION	STRONG	MODERATE	WEAK
See dictionary	1	2	3

C) CONFOUNDERS

(Q1) Were there important differences between groups prior to the intervention?

1 Yes
2 No
3 Can't tell

The following are examples of confounders:

1 Race
2 Sex
3 Marital status/family
4 Age
5 SES (income or class)
6 Education
7 Health status
8 Pre-intervention score on outcome measure

(Q2) If yes, indicate the percentage of relevant confounders that were controlled (either in the design (e.g. stratification, matching) or analysis)?

1 80 – 100% (most)
2 60 – 79% (some)
3 Less than 60% (few or none)
4 Can't Tell

RATE THIS SECTION	STRONG	MODERATE	WEAK
See dictionary	1	2	3

D) BLINDING

(Q1) Was (were) the outcome assessor(s) aware of the intervention or exposure status of participants?

1 Yes
2 No
3 Can't tell

(Q2) Were the study participants aware of the research question?

1 Yes
2 No
3 Can't tell

RATE THIS SECTION	STRONG	MODERATE	WEAK
See dictionary	1	2	3

E) DATA COLLECTION METHODS

(Q1) Were data collection tools shown to be valid?

1 Yes
2 No
3 Can't tell

(Q2) Were data collection tools shown to be reliable?

1 Yes
2 No
3 Can't tell

RATE THIS SECTION	STRONG	MODERATE	WEAK
See dictionary	1	2	3

GLOBAL RATING

COMPONENT RATINGS
Please transcribe the information from the gray boxes on pages 1-4 onto this page. See dictionary on how to rate this section.

A	**SELECTION BIAS**	**STRONG**	**MODERATE**	**WEAK**	
		1	2	3	
B	**STUDY DESIGN**	**STRONG**	**MODERATE**	**WEAK**	
		1	2	3	
C	**CONFOUNDERS**	**STRONG**	**MODERATE**	**WEAK**	
		1	2	3	
D	**BLINDING**	**STRONG**	**MODERATE**	**WEAK**	
		1	2	3	
E	**DATA COLLECTION METHOD**	**STRONG**	**MODERATE**	**WEAK**	
		1	2	3	
F	**WITHDRAWALS AND DROPOUTS**	**STRONG**	**MODERATE**	**WEAK**	
		1	2	3	Not Applicable

GLOBAL RATING FOR THIS PAPER (circle one):

1	STRONG	(no WEAK ratings)
2	MODERATE	(one WEAK rating)
3	WEAK	(two or more WEAK ratings)

With both reviewers discussing the ratings:

Is there a discrepancy between the two reviewers with respect to the component (A-F) ratings?

No Yes

If yes, indicate the reason for the discrepancy

1	Oversight
2	Differences in interpretation of criteria
3	Differences in interpretation of study

Final decision of both reviewers (circle one):

1	**STRONG**
2	**MODERATE**
3	**WEAK**

Appendix 7. NSC Criteria for appraising the viability, effectiveness and appropriateness of a screening programme

1. The condition

1. The condition should be an important health problem as judged by its frequency and/or severity. The epidemiology, incidence, prevalence and natural history of the condition should be understood, including development from latent to declared disease and/or there should be robust evidence about the association between the risk or disease marker and serious or treatable disease.

2. All the cost-effective primary prevention interventions should have been implemented as far as practicable.

3. 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.

2. The test

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

5. The distribution of test values in the target population should be known and a suitable cut-off level defined and agreed.

6. The test, from sample collection to delivery of results, should be acceptable to the target population.

7. There should be an agreed policy on the further diagnostic investigation of individuals with a positive test result and on the choices available to those individuals.

8. If the test is for a particular mutation or set of genetic variants the method for their selection and the means through which these will be kept under review in the programme should be clearly set out.

3. The intervention

9. There should be an effective intervention for patients identified through screening, with evidence that intervention at a pre-symptomatic phase leads to better outcomes for the screened individual compared with usual care. Evidence relating to wider benefits of screening, for example those relating to family members, should be taken into account where available. However, where there is no prospect of benefit for the individual screened then the screening programme shouldn't be further considered.

10. There should be agreed evidence based policies covering which individuals should be offered interventions and the appropriate intervention to be offered.

4. The screening programme

11. There should be evidence from high quality randomised controlled trials that the screening programme is effective in reducing mortality or morbidity. Where screening is aimed solely at providing information to allow the person being screened to make an

"informed choice" (such as Down's syndrome or cystic fibrosis carrier screening), there must be evidence from high quality trials that the test accurately measures risk. The information that is provided about the test and its outcome must be of value and readily understood by the individual being screened.

12. 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.

13. The benefit gained by individuals from the screening programme should outweigh any harms for example from overdiagnosis, overtreatment, false positives, false reassurance, uncertain findings and complications.

14. The opportunity cost of the screening programme (including testing, diagnosis and treatment, administration, training and quality assurance) should be economically balanced in relation to expenditure on medical care as a whole (value for money). Assessment against this criteria should have regard to evidence from cost benefit and/or cost effectiveness analyses and have regard to the effective use of available resource.

5. Implementation criteria

15. Clinical management of the condition and patient outcomes should be optimised in all health care providers prior to participation in a screening programme.

16. All other options for managing the condition should have been considered (such as improving treatment or providing other services), to ensure that no more cost effective intervention could be introduced or current interventions increased within the resources available.

17. There should be a plan for managing and monitoring the screening programme and an agreed set of quality assurance standards.

18. Adequate staffing and facilities for testing, diagnosis, treatment and programme management should be available prior to the commencement of the screening programme.

19. Evidence-based information, explaining the purpose and potential consequences of screening, investigation and preventative intervention or treatment, should be made available to potential participants to assist them in making an informed choice.

20. Public pressure for widening the eligibility criteria for reducing the screening interval, and for increasing the sensitivity of the testing process, should be anticipated. Decisions about these parameters should be scientifically justifiable to the public.

Appendix 8. TYR1 incidence in 51 U.S. states from 1st January 2001 to 31st December 2010[12]

State	Start TYR1-screening	Births	Cases	Incidence	Comments
Connecticut	1st May 2004	273,897	0	NA	
Maine	1st January 2001	135,961	0	NA	
Massachusetts	1st January 2001	785,953	2	*1:392,977*	
New Hampshire	1st July 2007	27,230	0	NA	
Rhode Island	1st July 2006	56,665	1	*1:56,665*	
Vermont	1st January 2003	48,316	0	NA	
Delaware	1st January 2003	97,626	0	NA	
District of Columbia	1st January 2006	29,416	1	*1:29,416*	No data available for 2008-2010; data not validated by the program.
Maryland	1st January 2001	718,032	2	*1:359,016*	
New Jersey	1st January 2005	656,334	2	*1:328,167*	
New York	1st January 2004	1,755,287	3	*1:585,096*	
Pennsylvania	1st July 2009	215,616	1	*1:215,616*	
Virginia	1st March 2006	504,757	0	NA	
West Virginia	1st February 2009	40,279	0	NA	Data not validated by the program.
Alabama	25th October 2004	378,676	1	*1:378,676*	
Florida	17th December 2009	233,429	1	*1:233,429*	
Georgia	1st January 2001	1,417,732	1	*1:1,417,732*	
Louisiana	1st January 2006	322,531	1	*1:322,531*	
Mississippi	1st June 2003	322,489	2	*1:161,245*	
North Carolina	1st January 2001	1,245,716	1	*1:1,245,716*	
South Carolina	1st November 2004	358,022	0	NA	
Tennessee	1st January 2004	615,964	2	*1:307,982*	
Illinois	1st January 2002	1,567,305	1	*1:1,567,305*	
Indiana	1st January 2003	704,310	0	NA	
Kentucky	1st January 2006	273,250	1	*1:273,250*	
Michigan	18th April 2005	690,037	0	NA	
Minnesota	1st January 2001	705,026	0	NA	
Ohio	1st August 2004	948,567	0	NA	
Wisconsin	1st March 2003	548,252	1	*1:548,252*	
Arkansas	1st July 2008	96,056	0	NA	
Iowa	1st January 2001	391,943	0	NA	Data not validated by the program.
Kansas	1st July 2008	105,394	0	NA	
Missouri	1st January 2005	483,977	0	NA	
Nebraska	1st July 2003	200,373	0	NA	
North Dakota	1st August 2004	64,625	0	NA	
Oklahoma	1st October 2008	119,427	0	NA	
South Dakota	1st January 2003	97,956	0	NA	
Arizona	5th April 2006	462,653	0	NA	
Colorado	1st January 2006	313,189	0	NA	
Montana	1st January 2004	59,836	0	NA	No data available for 2007-2008; data not validated by the program.
Nevada	1st January 2003	297,539	0	NA	
New Mexico	1st January 2007	114,820	0	NA	
Texas	6th December 2006	1,661,279	0	NA	
Utah	1st January 2006	276,174	1	*1:276,174*	
Wyoming	1st July 2006	32,458	0	NA	
Alaska	1st January 2003	86,374	1	*1:86,374*	
California	11th July 2005	2,997,046	10	*1:299,705*	
Hawaii	1st January 2003	149,783	0	NA	
Idaho	1st January 2003	184,644	0	NA	
Oregon	1st January 2002	428,110	0	NA	
Washington	1st July 2008	220,866	0	NA	

Appendix 9. Study characteristics and TMS screening methodology for tyrosinaemia type 1.

Study	Country, time period	Study design	Source and type of material	Age at specimen collection	Samples pooled?	Method of extraction & derivatisation	Type of TMS / TMS conditions	Analyte and cut-off	Positive screening results / Diagnostic confirmation	Narrative description
Allard 2004[33]	USA (Canada) New England Newborn Screening Programme Time period NR	Case-control study: Stored original newborn DBS specimens of 3 known TYR1 cases and 3,199 DBS of unaffected newborns.	Original newborn screening filter paper cards Controls: stored for up to 5 days at RT; Cases: stored at -20°C for 4-22 months. 3.2 mm (1/8 inch) diameter filter paper disc punch.	Controls: median 1.9 days; Cases: median 2 days, range 1-3 days.	No	Extraction of SUAC from residual DBS (already extracted with methanol for AA and AC analysis) with acetonitrile: water (80:20 by volume) containing 0.1% formic acid, 15 mmol/l hydrazine hydrate (0.1% by volume), and 100 nmol/l DOA as internal standard. Separate TMS analysis of SUAC-hydrazone.	Quattro LC triple-quadrupole tandem mass spectrometer (Micromass Inc, USA)/ Positive ion mode; Cone energy 20 V; Collision energy 11eV. SRM mode: SUAC-hydrazone *m/z* 155.1 → 137.1; DOA *m/z* 169.1 → 151.1.	Normal range SUAC < 2 µmol/l (study-derived)	NR / Clinically diagnosed cases, NR for controls	Allard et al. (2006) developed a method to extract SUAC from residual DBS following methanol extraction of markers for routine TMS analysis of newborn screening. SUAC was extracted using acetonitrile:water (80:20 by volume) containing 0.1% formic acid and 15 mmol/l hydrazine hydrate (0.1% by volume). SUAC-hydrazone was analysed in a separate TMS run. The analysis of stored original newborn DBS specimens of three clinically diagnosed TYR1 cases and 3,199 residual DBS of unaffected newborns used by the New England Newborn Screening Programme (analysed for SUAC within five days of AA/AC extraction) showed a clear discrimination between affected and normal newborns. Allard et al. concluded that using a SUAC cut-off of 2 µmol/l could result in a sensitivity and specificity of up to 100%. The reference standard used to confirm the absence of TYR1 in healthy controls was not reported.

Dhillon 2011[35]	California/ USA California Newborn Blood Spot Screening Programme 1 month	Case-control study: >1,000 NBS specimens identified as normal and stored NBS specimens of 6 confirmed TYR1 cases.	3.2 mm DBS punch	NR (newborn screening programme)	No	Simultaneous extraction of AA, AC, and SUAC from DBS using acetonitrile:water (8:2 by volume) containing 0.05% formic acid, 3.0 mmol/l hydrazine hydrate and $^{13}C_5$-SUAC as internal standard. Derivatisation of AA, AC and SUAC-hydrazone using butanolic-HCl. LC-TMS analysis of butyl esters in a single run.	Triple quadrupole tandem mass spectrometer, Micromass Quattro Micro (Waters Corporation) / Positive ion mode; Capillary voltage 3.2 kV; Multiple reaction monitoring; Butyl ester of SUAC-hydrazone: *m/z* 211.20 → 137.15; Butyl ester of $^{13}C_5$-SUAC hydrazone *m/z* 216.20 → 142.15	Cut-off SUAC 3 µmol/l (study-derived)	NR / NR	Dhillon et al. (2011) described a simultaneous extraction of AA, AC, and SUAC from DBS using acetonitrile:water (8:2 by volume) containing 0.05% formic acid and 3.0 mmol/l hydrazine hydrate. Liquid chromatography (LC)-TMS analysis of butyl esters was performed in a single run. The analysis of more than 1,000 NBS specimens identified as normal and stored NBS specimens of six confirmed TYR1 cases identified by the California NBS programme found a clear-cut demarcation between normal and affected newborns. Using a study-derived cut-off of 3 µmol/l identified all specimens analyzed correctly. The reference standard used to confirm the absence of TYR1 in healthy controls was not reported.

Study	Country, time period	Study design	Source and type of material	Age at specimen collection	Samples pooled?	Method of extraction & derivatisation	Type of TMS / TMS conditions	Analyte and cut-off	Positive screening results / Diagnostic confirmation	Narrative description
La Marca 2008[36]	Italy January 2007 to May 2007	13,000 newborn screening spots from healthy controls and 10 stored DBS samples from 6 confirmed TYR1 cases.	DBS using heel stick, spotted on filter paper (903, Whatman), 3.2 mm punch (3.4 µl blood) used.	Controls: 48-72 h of life; Cases: 3 days-11 months.	No	Simultaneous extraction of AA, AC and SUAC: Addition of DOA (or $^{13}C_2$-SUAC) as internal standard to the methanolic solution of deuterated AC and AA; Extraction and derivatisation of SUAC in a single step using 3 mmol/l hydrazine in water/methanol (50:50); Butylation; Simultaneous TMS measurement of AC, AA and SUAC-hydrazone as butyl esters.	Applied Biosystems/MDS Sciex API 4000™ triple-quadrupole MS equipped with a TurboV-Spray® source with turbo gas temperature set at 425ºC / Positive ionisation polarity +5500 V. Multiple reaction monitoring: Butyl ester of SUAC-hydrazone *m/z* 211 → 137; Internal standards Butyl ester of $^{13}C_2$-SUAC-hydrazone *m/z* 219 → 139; Butyl ester of DOA *m/z* 225 → 151. Declustering enery 55V; Collision energy 19eV.	SUAC Normal range < 2.4 µmol/l (median+5SD)	NR / NR	La Marca et al. (2008) developed a simultaneous extraction of AA, AC and SUAC using a solution of 3 mmol/l hydrazine in water:methanol (50:50). AC, AA and SUAC-hydrazone were analysed as butyl esters in a single TMS run. Using 13,000 prospectively collected DBS from the Tuscan Newborn Screening Programme as healthy controls and 10 stored DBS from 6 confirmed TYR1 cases, La Marca et al. found a clear-cut demarcation between normal and affected children. Using a cut-off of 2.4 µmol/l distinguished true positives from controls, false positives and false negatives. The reference standard used to confirm the absence of TYR1 in healthy controls was not reported.

Study	Country, time period	Study design	Source and type of material	Age at specimen collection	Samples pooled?	Method of extraction & derivatisation	Type of TMS / TMS conditions	Analyte and cut-off	Positive screening results / Diagnostic confirmation	Narrative description
La Marca 2011[30]	Tuscany/ Italy January 2007-2010	136,075 screened in Tuscan expanded newborn screening programme; Overlap of 13,000 samples reported by La Marca 2008.	DBS	NR (48-72 h of life, from La Marca 2008[36])	No	NR (see La Marca 2008[36])	NR (see La Marca 2008[36])	SUAC Normal value < 2 µmol/l	NR / Detection of SUAC in urine and plasma	The approach described by La Marca et al. 2008[36] was used for prospective newborn screening in 136,075 newborns born in Tuscany (Italy) since January 2007 for about 3 or 4 years. Two affected children were identified using a cut-off of 2 µmol/l with no false positive screening result, resulting in a PPV of 100%. Follow-up of 136,073 screen-negatives was not described, the duration and methods of follow-up (if performed) as well as losses to follow-up are unclear. La Marca et al. reported no known false negative result and 100% sensitivity.
Lund 2012[16]	Denmark, Faroe Islands, Greenland Routine expanded newborn screening programme February 2009 to March 2011 (26 months)	Prospective routine expanded newborn screening in 140,565 newborns	DBS using heel prick, spotted om filter paper (Schleich and Schuell 903 filter paper until 2010, then gradually replaced by the Ahlstrom 226)	2-3 days; Median 2.5 days. Preterm newborns: repeated test at gestational age 32 weeks or when oral feeding had been established.	No	PerkinElmer Neobase non-derivatized MS/MS kit™ (3040-0010)	Waters Micromass Quattro micro™ tandem mass spectrometer / NR	SUAC > 2.1 U	Flagged DBS samples re-analysed in duplicates. If the abnormal profiles were reproduced, referral to Center for Inherited Metabolic Disorders, Copenhagen University Hospital / Urine organic acids, plasma amino acids, molecular genetic analyses	Lund et al. (2012) reported the results of routine expanded newborn screening in Denmark, the Faroe Islands, and Greenland. Screening for TYR1 was introduced in February 2009 using a commercially available mass spectrometry kit (PerkinElmer Neobase non-derivatized MS/MS kitTM [3040-0010]). Prospective screening of 140,565 newborns over 26 months detected one true positive case of TYR1 with no false positive result (PPV 100%). The methods and duration of follow-up for those who screened negative as well as losses to follow-up were not reported. All children with suspected inborn errors of metabolism in Denmark, the Faroe Islands, and Greenland are diagnosed and treated in the same centre (Copenhagen University Hospital). No false negative screening result was known to the authors at the time of writing (about 14 months after study period).

Study	Country, time period	Study design	Source and type of material	Age at specimen collection	Samples pooled?	Method of extraction & derivatisation	Type of TMS / TMS conditions	Analyte and cut-off	Positive screening results / Diagnostic confirmation	Narrative description
Metz 2012[31]	Austria Austrian Newborn Screening Programme 1 month	Prospective routine newborn screening in 4,683 consecutive newborns and stored DBS samples from 3 confirmed TYR1 cases.	DBS (Ahlstrom226 Paper, ID Biological, SC), one 3.2 mm punch	NR	No	MassChrom® Amino Acids and Acylcarnitines from Dried Blood; Chromsystems, Munich/Germany: Separate SUAC extraction (including $^{13}C_5$-SUAC as internal standard) from residual DBS after extraction of AA and AC; Derivatisation solution with ~0.0005% hydrazine derived reagent; Transfer of SUAC-hydrazone to AA and AC residues; Butylation; Simultaneous TMS measurement of AC, AA and SUAC-hydrazone as butyl esters.	FIA-TMS analyses on a certified TQ-Detector MS system for newborn screening (Waters, Milford, MA, USA) / Positive ion mode using a dwell time of 0.05 s. Butyl ester of SUAC-hydrazone *m/z* 211 → 109. Butyl ester of $^{13}C_5$-SUAC-hydrazone *m/z* 216 → 114. Cone (V): 20 Collision (eV): 24	SUAC Preliminary cut-off 1.29 µmol/l; Derived from first 4,000 specimens of unaffected newborns born after 32 weeks of gestation and samples not obtained within 36 h after birth.	DBSs from potentially affected newborns were re-tested at least in duplicates from 2 separate blood spots from the same DBS card. / In case of positivity, diagnostically confirmed in accordance with institutional guidelines.	Metz et al. (2012) evaluated another commercially available mass spectrometry kit (MassChrom® Amino Acids and Acylcarnitines from Dried Blood; Chromsystems, Munich/Germany) in a one-month study period during the national routine Austrian Newborn Screening Programme. DBS specimens from 4,683 newborns were collected consecutively; no true positive or false positive results were obtained with a preliminary SUAC cut-off of 1.29 µmol/l. Analysis of stored DBS cards correctly identified 3 known patients with TYR1 (PPV 100%). Follow-up of screen negatives was not defined and losses to follow-up were not reported.

Study	Country, time period	Study design	Source and type of material	Age at specimen collection	Samples pooled?	Method of extraction & derivatisation	Type of TMS / TMS conditions	Analyte and cut-off	Positive screening results / Diagnostic confirmation	Narrative description
Morrissey 2011[20]	New York State/ USA December 2007-? (2008 and 2009, over 24 months)	~500,000 newborns screened prospectively in New York State newborn screening programme	DBS, 3.2 mm punch from a Guthrie card (~3.1µl blood)	NR	No	Extraction of SUAC from residual DBS (after methanol extraction of AA and AC) after overnight drying using acetonitrile:water (80:20, containing 0.1% formic acid, 0.1% hydrazine, plus $^{13}C_5$-SUAC as internal standard), TMS analysis of SUAC hydrazone one day after AA/AC analysis.	Two Waters Corp Micro LC TMS (Manchester, UK) with Hewlett-Packard/Agilent Technologies series 1100 HPLC pumps. TQD TMS and Acquity UPLC system (Waters Corp) for handling overload and maintenance / Selected ion monitoring: SUAC-hydrazone 155.1 → 137.1; $^{13}C_5$-SUAC-hydrazone 160.1 → 142.1.	SUAC ≥ 3.00 µmol/l for retest; Average (initial and retest) 3.00-5.00 µmol/l: repeat DBS specimen requested; Average (initial and retest) ≥ 5.00 µmol/l: Immediate referral.	Initial SUAC ≥ 3.00 µmol/l retested in duplicate. Average (initial and retest) SUAC 3.00-5.00 µmol/l, repeat DBS request. Average SUAC ≥ 5.00 µmol/l immediate referral to the appropriate specialty care center./ Prenatal testing or plasma AA and SUAC with or without urine organic acids and liver function.	Morrissey et al. (2011) reported data for over 24 months of prospective newborn screening in New York State. SUAC was extracted from residual DBS (following methanol extraction of amino acids and acylcarnitines after overnight drying using acetonitrile:water (80:20, containing 0.1% formic acid and 0.1% hydrazine). Any sample with initial SUAC ≥ 3.00 µmol/l was retested in duplicate. For any specimen with an average SUAC value (initial and retest) between 3.00 and 5.00 µmol/l a repeat blood spot specimen was requested. For any sample with an average SUAC value of ≥5.00 µmol/l an immediate referral was made to the appropriate specialty care centre. There were 5 screen positive results among approximately 500,000 samples screened: 2 of these were in the range 3.00-5.00 µmol/l, and a repeat specimen screened negative. Of 3 samples with SUAC ≥5.00 µmol/l, 2 patients were diagnosed with TYR1 (true positives), the other patient was a false positive (PPV 67%). The exact number and timing of follow-up for screen-negatives were not reported; losses to follow-up are unclear.

Study	Country, time period	Study design	Source and type of material	Age at specimen collection	Samples pooled?	Method of extraction & derivatisation	Type of TMS / TMS conditions	Analyte and cut-off	Positive screening results / Diagnostic confirmation	Narrative description
Sander 2006[27]	Germany 16 weeks	Prospective newborn screening study in 61,344 unselected newborns plus retrospective analysis of stored original DBS cards from 2 confirmed TYR1 patients.	DBS on S&S 903 filter paper (3.2 mm)	36 – 72 hours after birth	No	Extraction of SUAC from residual DBS (already extracted with absolute methanol for AC and AA analysis) using acetonitrile-water (80:20 by volume) containing formic acid, 15 mmol/l hydrazine hydrate, and unlabelled DOA as internal standard. TMS analysis of SUAC hydrazone in a separate run.	MS/MS micro™ and Quatro LC™; Waters/Micromass Inc. / Positive ion mode, cone energy 20 V, collision voltage 10 eV, dwell time at 9.1 s; Multiple-reaction monitoring mode; SUAC-hydrazone *m/z* 155.2 → 137.1 and *m/z* 155.2 → 109.1; DOA *m/z* 169.3 → 151.2.	SUAC > 10 µmol/l	NR / Urinary SUAC and phenolic acids or tyrosine metabolites.	Sander et al. (2006) presented their experience with TYR1 screening in a 16-week prospective screening study in Germany. They used the method described by Allard et al. (2004)[33] and extracted SUAC from residual DBS that had already been extracted for AA and AC using methanol. Using a cut-off of 10 µmol/l, the analysis of 61,344 unselected newborn DBS samples identified 2 true positive TYR1 cases with no false positive results (PPV 100%). Diagnostic sensitivity has not been evaluated with follow-up of all 61,344 newborns for TYR1. Losses to follow-up of those who screened negative were not reported. No false negative results have been reported at time of writing. Retrospective analysis of stored neonatal screening samples of two additional known TYR1 patients revealed increased SUAC concentrations of 46 and 169 µmol/l, respectively.

Study	Country, time period	Study design	Source and type of material	Age at specimen collection	Samples pooled?	Method of extraction & derivatisation	Type of TMS / TMS conditions	Analyte and cut-off	Positive screening results / Diagnostic confirmation	Narrative description
Turgeon 2008[34]	Minnesota/ USA Mayo Clinic's supplemental newborn screening programme Time period NR	Case-control validation study in 13,521 stored random newborn screening samples not suggestive of TYR1, based on 2nd-tier screening[39] and 11 stored original DBS from confirmed TYR1.	DBS	NR	No	Parallel extraction of SUAC from residual DBS (already extracted with methanol for AA and AC analysis) using acetonitrile/ water/formic acid solution (80:20:0.1, v/v/v) containing 0.1% hydrazine monohydrate (15 mmol/l) and $^{13}C_5$-SUAC as internal standard. Combined TMS analysis of AC and AA butyl esters, SUAC hydrazone and $^{13}C_5$-SUAC hydrazone.	Triple-quadrupole MS/MS (Applied Biosystems/MDS Sciex API 3000) / Positive ion mode (source voltage, 5500 V). Method optimisation for detection of SUAC by SRM: SUAC-hydrazone *m/z* 155.0 → 137.0; $^{13}C_5$-SUAC-hydrazone *m/z* 160.0 → 142.0.	SUAC > 5.0 µmol/l	NR / NR for cases, 2nd-tier screening approach for controls.	Turgeon et al. (2008) reported a method in which SUAC was extracted from residual DBS (extracted for AA and AC using methanol) using acetonitrile:water:formic acid solution (80:20:0.1, v:v:v) containing 0.1% hydrazine monohydrate (15 mmol/l). Extracts are combined and TMS analysis of AC and AA as butyl esters and SUAC-hydrazone was performed in a single run. In a validation study, 13,521 stored random NBS samples from Mayo Clinic's Supplemental Newborn Screening programme not suggestive of TYR1 based on a 2nd-tier screening approach by Magera et al. (2006)[39] and 11 stored original DBS from confirmed TYR1 cases were analysed. Setting the cut-off for SUAC at 5.0 µmol/l allowed clear discrimination of the control population from TYR1 patients.

Study	Country, time period	Study design	Source and type of material	Age at specimen collection	Samples pooled?	Method of extraction & derivatisation	Type of TMS / TMS conditions	Analyte and cut-off	Positive screening results / Diagnostic confirmation	Narrative description
Zytkovicz 2013[32]	Massachusetts/USA New England Newborn Screening Programme 1 June 2008 to 30 June 2012 (4 years 1 month)	518,687 samples received in New England newborn screening programme (491,472 [94.8%] born nationally; 27,215 [5.2%] born internationally) plus 3 stored DBS samples from 2 confirmed TYR1 cases analysed using quantitative SUAC assay only	DBS (1/8 inch punch)	99.4% newborn period (less than 1 month); 0.6% over 1 month of age.	Yes (up to 8 samples pooled)	Pooled sample assay: SUAC from residual DBS (previously extracted with methanol for AA and AC) was extracted using acetonitrile: water: formic acid (80:20:0.1%) containing 17.7 mM hydrazine and 0.4 µM $^{13}C_5$-SUAC as internal standard. Up to 8 sample extracts pooled; TMS analysis of SUAC hydrazone. Quantitative assay: Untreated (newly punched) DBS were extracted and analysed as above but not pooled.	Waters Quattro micro MS/MS / SUAC-hydrazone *m/z* 155.2 → 137.1 and *m/z* 155.2 → 109.1; $^{13}C_5$-SUAC-hydrazone *m/z* 160.2 → 142.1 and *m/z* 160.2 → 114.1.	Pooled assay: SUAC > 0.55 µM re-analysed individually. Quantitative assay: SUAC > 4 µM (recently reduced to 3.3 µM) positive; SUAC 1.0-3.3 µM intermediate.	Pooled assay: SUAC > 0.55 µM re-analysed individually. Quantitative assay: Samples with SUAC 1.0-3.3 µM → repeat DBS specimen. SUAC > 4 µM (recently reduced to 3.3 µM) → diagnostic testing recommended / NR (1 case had symptoms consistent with TYR1)	Zytkovicz et al. (2013) developed a TMS assay to screen for TYR1 from DBS using pooled extracts to increase high throughput screening. The method described by Allard et al. (2004)[33] was modified: SUAC from residual DBS was extracted, up to 8 sample extracts were pooled and the SUAC-hydrazone derivative was analysed by TMS. Pooled samples yielding SUAC levels greater than 0.55 µM were re-analysed individually. For the individual assay, SUAC > 4 µM were considered positive (this cut-off was later reduced to 3.3 µM) and samples with SUAC between 1.0 and 3.3 µM were considered indeterminate and required a repeat DBS specimen. The method was used prospectively during New England Newborn Screening Programme over 4 years 1 month in 518,687 samples. Five cases screened positive, 3 of these have been confirmed to have TYR1 and follow-up information on the remaining 2 born out of the country was not available. Four newborns had indeterminate SUAC levels between 1.0 and 3.3 µM. A repeat screen was received on only one, and the result was negative. The other 3 newborns were born out of the country and no repeat DBS sample was received. Follow-up of screen-negatives was not described and losses to follow-up are unclear; time period between end of the study and time of manuscript submission was less than 3 months.

AA, amino acids; AC, acylcarnitines; DBS, dried blood spot; DNA, deoxyribonucleic acid; DOA, 5,7-dioxooctanoic acid; FIA, flow injection analysis; IEM, inborn errors of metabolism; LC, liquid chromatography; m/z; mass-to-charge ratio; NBS, newborn blood spot screening; NR, not reported; RT, room temperature; SUAC, succinylacetone; SRM, selected reaction monitoring; TMS, tandem mass spectrometry or tandem mass spectrometer; TYR1, Tyrosinaemia type 1.

Appendix 10. Study quality of included studies (key question 2) according to untailored QUADAS-2 (20% checked)[4]

Study	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patients	Index test	Reference standard
Allard 2004[33]	High	High	Unclear	Unclear	High	Low	Unclear
Dhillon 2011[35]	High	High	Unclear	Unclear	Unclear	Low	Unclear
La Marca 2008[36]	High	Unclear	Unclear	High	High	Low	Unclear
La Marca 2011[30]	Low	Unclear	Unclear	High	High	Low	Low
Lund 2012[16]	Low	Unclear	Unclear	High	High	Low	Low
Metz 2012[31]	Low	Unclear	Unclear	High	Unclear	Low	Unclear
Morrissey 2011[20]	Low	Low	Unclear	High	Unclear	Unclear	Low
Sander 2006[27]	Low	Unclear	Unclear	High	High	Low	Low
Turgeon 2008[34]	High	High	High	High	Unclear	Low	Unclear
Zytkovicz 2013[32]	Low	Unclear	Unclear	High	High	Unclear	Unclear

Appendix 11. Study quality of included studies (key question 2) according to adjusted QUADAS-2 (100% checked)[4]

Study	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patients	Index test	Reference standard
Allard 2004[33]	High	High	Unclear	High	High	Low	Unclear
Dhillon 2011[35]	High	High	Unclear	High	Unclear	Low	Unclear
La Marca 2008[36]	High	High	Unclear	High	High	Low	Unclear
La Marca 2011[30]	Unclear	Unclear	High	High	High	Low	High
Lund 2012[16]	Low	Unclear	High	High	High	Low	High
Metz 2012[31]	Unclear/High†	High	High	High	Unclear	Low	High
Morrissey 2011[20]	Unclear	Low	High	High	Unclear	Low	High
Sander 2006[27]	Unclear	Unclear	High	High	High	Low	High
Turgeon 2008[34]	High	High	High	High	High	Low	High
Zytkovicz 2013[32]	High	Unclear	High	High	High	Low	High

† when stored dried blood spot specimens from confirmed TYR1 cases were included (case-control analysis).

Appendix 12. Study level description of included studies for key question 3 (early vs. late treatment).

Study	Narrative description
Québec study	
Larochelle 2012[40]	Larochelle et al. (2012) compared the outcome of children with TYR1 born during the first 10 years that Nitisinone was available in Québec with that of patients born in the preceding decade. Seventy-eight TYR1 patients with confirmed diagnosis by the presence of elevated levels of SUAC in blood or urine born between February 1984 and February 2004 (20 years) in Québec were included in this study. Patients were divided in three groups: never treated with Nitisinone (n=28); treated after 30 days of age (late-NTBC, n=26) and treated on or before 30 days of age (early-NTBC, n=24). The clinical course of patients was recorded until orthotopic liver transplantation (OLT), death, or until 1st August 2009. Ten patients (36%) who did not receive Nitisinone died (8 before OLT and 2 after OLT) while two deaths after OLT (8%) occurred in the late-NTBC group ($p<0.01$ vs never-treated) and no deaths in the early-NTBC group ($p<0.001$ vs never-NTBC; $p<0.05$ vs late-NTBC). OLT was performed in 20 never treated patients (71%) at a median age of 26 months, while seven late-treated patients (27%, $p<0.001$ vs never treated) and no early-treated patient ($p<0.001$ vs never-NTBC; $p<0.001$ vs late-NTBC) required OLT. No hospitalisations for acute complications of TYR1 (neurological crisis included) occurred during 2,593 patient months in the early-treated group. Acute complications occurred during 43 of 3,673 patient months (1.2%) in the late-treated group; no acute events occurred after the first dose of Nitisinone. In the never-treated group, TYR1-related hospitalisations occurred in 141 of 777 patient month (18.1%, $p<0.001$ vs early-treated; $p<0.001$ vs late-treated). However, in the late treated group eight patients had a tyrosinemia diagnosis at least 2 years before receiving nitisinone due to Nitisinone only becoming available in 1992 so these patients have limited applicability to the UK situation where treatment would be administered upon diagnosis, and a further three were either not detected by screening or moved into the area without screening (it is not clear which) so if not detectable by screening these patients would not benefit from screening. None of the remaining patients in the late treatment group experienced transplant or death in around 10 years of follow-up, indicating that whilst Nitisinone is an effective treatment, this study does not provide evidence that early treatment following screen detection is more effective than late treatment following symptomatic detection.
Simoncelli 2015[41]	Simoncelli et al. (2015) performed a cost-consequence analysis of Nitisinone for treatment of TYR1. This study included 95 children who were treated in Québec between January 1, 1984, and January 1, 2009 (25 years) with a diagnosis of TYR1 confirmed by neonatal genetic screening. Three groups were designated as described by Larochelle and others[40] (see above row), except that children born between February 1, 2004 and January 1, 2009 were also included in the early-NTBC group (n=41). NTBC treatment was associated with a significant decline in hospital admissions, in terms of both number of admissions per person-year (0.83, 0.41, and 0.16 for never-NTBC, late-NTBC, and early-NTBC groups, respectively; $p<0.001$) and length of stay (7.6, 3.2, and 0.4 days respectively); $p<0.001$). NTBC therapy was associated with a reduction of porphyria-like neurological crises: 51 crises affected 14 (50%) patients in the never-NTBC group, 16 crises affected 5 (19%) patients in the late-NTBC group (all occurring before NTBC treatment was initiated), whereas no crises occurred in the early-NTBC group ($p<0.001$). OLT was performed in 20 patients (71%) in the never-NTBC group, while seven late-treated patients (27%) and no early-treated patient required OLT ($p<0.001$). Ten never NTBC treated patients (36%) died while two deaths (8%) occurred in the late-NTBC group and none in the early-NTBC group ($p<0.001$).

Study	Narrative description
Birmingham study	
Bartlett 2014[42]	Bartlett et al. (2014) presented their experience of the management of children with TYR1 in a single centre before and after the introduction of Nitisinone and how this affected outcomes and need for OLT. Thirty-eight patients who were treated for TYR1 at the Birmingham Children's Hospital (UK) from 1989-2009 were included in this study. Nitisinone treated patients (n=31) were subdivided into early (<2 months, n=11), intermediate (2-6 months, n=11) and late (>6 months, n=9) groups according to age at presentation. Of the seven patients never treated with NTBC, 6 (85.7%) required OLT while only 7 of the 31 (22.6%) NTBC-treated patients proceeded to OLT (p=0.004), which adds evidence that Nitisinone is an effective treatment. No patient in the early-NTBC group, 3/11 (27%) patients in the intermediate-NTBC group and 4/9 (44%) patients in the late-NTBC group subsequently underwent OLT. In the intermediate group all 3 cases were treated immediately after detection. In the late group 2 out of 4 cases were treated with a delay of 6 months after diagnosis, so are not applicable to the clinical question. This provides some evidence in a very small sample that treatment in the first 2 months may reduce the risk of liver transplant. Those patients requiring OLT started NTBC treatment at a median age of 428 (86-821) days (61 weeks) as compared to those who did not require OLT who started treatment at a median age of 52 (2-990) days (7.5 weeks) (p=0.004).However this analysis includes patients from before Nitisinone became available in 1992, and so will have limited applicability to the comparison of interest, screen detected vs symptomatically/cascade testing detected when nitisinone is administered immediately upon detection.
McKiernan 2015[43]	McKiernan et al. (2015) described the outcome of children with TYR1 treated with Nitisinone following cascade testing and differential diagnosis from current PKU screening. Cases detected in this way were compared their outcome with index cases who had presented clinically. Seventeen children with TYR1 from Birmingham Children's Hospital (UK) treated with Nitisinone following early detection (tested, n=12) or index cases who presented clinically (n=5) were included in this study. NTBC treatment was started in the tested group at a median of 4 (range 2-52) days. The other cases presented at a median of 4 (range 1.5-17) months. No death occurred in the tested group while 2 of 5 (40%) clinically presenting infants died (one was born prior NTBC availability and died prior OLT, the other was born at 25 weeks gestation and died from respiratory complications of prematurity). No patient in the tested group required OLT; all were reported as currently clinically normal with no clinical, biochemical or radiological evidence of liver or kidney disease. In clinically presenting cases, one failed to respond to Nitisinone and underwent OLT at 5 months and repeat OLT aged 15 for chronic rejection while the other two surviving patients were treated with diet and Nitisinone and were clinically stable with compensated liver disease.
Santra 2008[44]	Santra et al. (2008) documented the incidence and progression of renal tubular dysfunction in children with TYR1 treated with Nitisinone at Birmingham Children's Hospital (UK). Twenty-one patients with TYR1 treated with Nitisinone for at least 12 months were included. Children were classified according to the phenotype of liver disease at presentation: acute liver failure (n=9), chronic liver disease (n=7), or pre-clinical presentation (n=5). Median age at presentation was less than 1 week (range < 1 week to 2 weeks) in the pre-clinical group, 17 weeks (range 1 month to 2 years) in infants with acute liver failure and 60 weeks (range 2 months to 9 years) in infants with chronic liver and/or kidney disease. Follow-up time ranged from one year (100% of children) to 10 years (29% of children). All TYR1 patients had proteinuria at presentation with high values seen even in the children who were diagnosed pre-clinically. Children who presented with acute liver failure were more likely to be hypophosphataemic (p<0.01) and have excessive phosphaturia (p=0.05) than children who were diagnosed pre-clinically, while children who presented with chronic hepatic and/or renal dysfunction had intermediate values. Four of 16 (25%) children who presented clinically required phosphate supplementation whilst none of the children diagnosed pre-clinically did. After Nitisinone and dietary treatment were started, all three markers of tubular dysfunction normalised within one year and remained normal at follow-up of up to 10 years. No child redeveloped tubular dysfunction after starting on Nitisinone treatment.

Study	Narrative description
Multicentre survey	
Mayorandan 2014[45]	Mayorandan et al. (2014) collected cross-sectional data via questionnaires on diagnosis, management, monitoring and outcome of TYR1 in an international cohort. Twenty-one metabolic centres from Europe, Turkey and Israel contributed data of 168 TYR1 patients. One-hundred forty-eight patients (88%) with data on age at NTBC start and clinical course were divided into four groups, depending on initiation of NTBC treatment: <1 month (n=37); 1-6 months (n=45), 7-12 months (n=20) and >12 months (n=46). Early treatment with NTBC (<1 month) was associated with a lower rate of complications, especially hepatocellular carcinoma (HCC), requirement of OLT, liver cirrhosis, hepatomegaly, rickets and renal dysfunction compared to later NTBC treatment. In this study cases detected before 1992 were excluded, increasing applicability to the research question of interest in the UK. However, it is not clear what the relative contributions of timing of nitisinone treatment is to the effect, in comparison to confounders such as differences in disease spectrum and quality of healthcare received by timing of diagnosis and nitisinone administration in this large international survey.

NTBC, Nitisinone, Orfadin®; OLT, orthotopic liver transplantation; TYR1, tyrosinaemia type 1.

Appendix 13. Main differences between rapid review and systematic review results (systematic review methods applied to key question 2 only).

	Rapid review	Systematic review
Relevant clinical differences	• 10 studies included for key question 2. • 4 studies included for key question 1.	• No additional studies identified for key question 2. • 3 additional studies identified for key question 1 (Incidence) by broadened SUAC search.
QUADAS-2[4]	• Untailored QUADAS-2 without guidance notes; • 20% in duplicate; • Disagreements: 6/14 scores (43%); • "Unclear" ratings: 29/70 (41%); • "High" ratings: 22/70 (31%).	• Adjusted QUADAS-2 with guidance notes; • 100% in duplicate; • Disagreements: 20/70 scores (29%); • "Unclear" ratings: 16/70 (23%); • "High" ratings: 42/70 (60%).

SUAC, succinylacetone.