UK National Screening Committee

Newborn screening for galactosaemia

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

Version: 2

Author: Bazian

Date: 8th May 2020

The UK National Screening Committee secretariat is hosted by Public Health England.

About the UK National Screening Committee (UK NSC)

The UK NSC advises ministers and the NHS in the 4 UK countries about all aspects of population screening and supports implementation of screening programmes. Conditions are reviewed against evidence review criteria according to the UK NSC's evidence review process.

Read a complete list of UK NSC recommendations.

UK NSC, Floor 5, Wellington House, 133-155 Waterloo Road, London, SE1 8UG
www.gov.uk/uknsc
Twitter: @PHE_Screening Blog: phescreening.blog.gov.uk

For queries relating to this document, please contact: phe.screeninghelpdesk@nhs.net

Published Month 20XX

Contents

Plain English summary

Galactosaemia is a hereditary condition. A person with this condition cannot break down galactose, a sugar found in milk. Babies with galactosaemia have inherited a faulty gene which means that they cannot make an enzyme called GALT. This is the enzyme needed to break down galactose. The signs and symptoms of galactosaemia vary from mild to severe.

Babies consume galactose from birth, through breast milk or milk-based formula. Babies with galactosaemia cannot break down galactose, so it builds up in their bodies. High levels of galactose in the body can cause feeding difficulties, sickness and liver damage. It can also cause long-term complications such as speech difficulties and delayed development. Treatment involves avoiding foods that contain galactose for life. During infancy, babies consume soy-based formula instead of breast milk and milk-based formula.

Newborn screening for galactosaemia is not currently recommended in the UK. This was because the last review did not find evidence that screening can reduce the risk of complications for the baby. It also did not show a clear benefit from starting the milk-free diet earlier in the first days of life compared with later.

This review looked for new evidence published since the last review. It aimed to see whether:

- there is enough evidence to show the age at which symptoms develop
- screening tests are reliable enough to identify babies with galactosaemia
- screening improves short-term and long-term outcomes.

The balance between benefits and harms remains unclear. So, the UK National Screening Committee does not recommend screening babies for galactosaemia. This is because:

- babies show symptoms at around 7 days of age, before the screening results are available
- screening tests are not accurate enough and would misdiagnose many healthy babies
- it is unclear whether early treatment can improve long-term health outcomes.

Executive summary

Purpose of the review

This UK National Screening Committee (NSC) evidence summary aimed to assess whether there have been significant developments in the evidence base since the last UK NSC evidence summary of newborn screening for classic galactosaemia[1] published in 2015. The purpose of this evidence synthesis was to assess whether the current UK NSC recommendation, not to implement screening for classic galactosaemia, should be reconsidered.

Background

Galactosaemia is a hereditary disorder of galactose metabolism, a type of condition known as an inborn error of metabolism (IEM). It can be caused by deficiency in one of the 3 enzymes in the galactose metabolic pathway: galactokinase (GALK), galactose-1-phosphate uridyl transferase (GALT) and UDP-galactose 4- epimerase (GALE). The condition being reviewed here, and the potential candidate for newborn screening, is classic galactosaemia, caused by deficiency of the second enzyme in the pathway, GALT. This results in accumulation of galactose and its metabolites in the blood, which have toxic effects. Deficiencies of the other 2 enzymes, GALK and GALE, cause conditions that are biochemically and clinically distinct from classic galactosaemia,[2] and fall outside the scope of this evidence summary. Classic galactosaemia has an autosomal recessive pattern of inheritance, meaning that affected individuals have inherited 2 pathogenic variants (disease-causing mutations) of the *GALT* gene, one from each parent.

Classic galactosaemia has been estimated to affect 1 in 44,000 newborns in the UK[3] though the incidence will be affected by frequency of pathogenic variants in the population and factors such as consanguinity. In Ireland, the incidence has been estimated at 1 in 34,000 newborns among the non-Traveller population but 1 in 430 among the Traveller community.[4]

The International Clinical Guideline of Classic galactosaemia[5] defines the condition as:

- absent or profoundly deficient GALT enzyme activity in red blood cells (typically <1%)
- and/or DNA analysis demonstrating 2 pathogenic variants (alleles) of the *GALT* gene.

Individuals usually present in the first week of life with signs of liver dysfunction (such as jaundice and hepatomegaly), lethargy, vomiting, diarrhoea or sepsis.[2] Management of classic galactosaemia involves a galactose-restricted diet, therefore breastfeeding and

whey-based formula should be replaced with soy-based or elemental formula.[5] Avoidance of milk products needs to be continued for life, though even with galactose-restriction, long-term complications are common. These include neurodevelopmental problems, delayed puberty and fertility problems, reduced bone mineral density and cataracts.

Newborn screening for classic galactosaemia includes measuring total galactose (galactose plus the metabolite galactose-1-phosphate, Gal-1-P) and/or GALT enzyme activity in the newborn dried blood spots (DBS). However, the index test (single or combination), laboratory assay and cut-offs have all varied widely in screening pilots and programmes to date, and demonstrated different test performance. The value of screening may also be debated given the early symptomatic presentation and risk of long-term complications even with treatment. Additionally, screening may detect newborns with *GALT* variants (such as Duarte galactosaemia) that are characterised by enzyme activity above 10 to 15% and would not have caused symptoms. Hence, there is currently wide international variation in practice with screening performed in all US states, 8 of 15 Canadian territories, 14 of 20 countries in Latin America, 9 of 24 countries in Asia Pacific and 13 of 48 European countries.[6]

Focus of the review

This review aimed to evaluate the evidence on newborn screening for classic galactosaemia published since the 2015 UK NSC evidence summary.[1] Specifically, new evidence was collected to answer the following 3 key questions:

1. What is the median age of presentation of classic galactosaemia? (Criterion 1)
2. What is the accuracy of the available screening tests to detect classic galactosaemia? (Criteria 4 and 5)
3. Does early initiation of treatment for individuals with classic galactosaemia provide better short- and long-term outcomes? (Criterion 9)

A rapid review search for these questions was conducted in January 2020 for studies published since January 2014, the search date of the last UK NSC evidence summary.

Recommendation under review

The UK NSC does not currently recommend universal newborn screening for classic galactosaemia. This recommendation was made on the basis of the last UK NSC evidence summary on the topic, published in 2015.[1]

Prior to the last UK NSC evidence summary, a 1997 health technology assessment (HTA)[3] had concluded that a screening programme for classic galactosaemia could not be

supported in the UK. The authors considered that this was a well-defined disorder with known UK incidence and one that was known to be associated with significant morbidity. However, it was unclear whether there was a safe, simple and robust screening test and uncertainty whether earlier treatment could improve long-term outcomes.[3]

The 2015 UK NSC evidence summary had assessed the evidence published since the 1997 HTA[3] and focused on key questions relating to the HTA's conclusions. The review concluded that:

- while classic galactosaemia was sufficiently understood, a screening programme may identify individuals carrying variants of uncertain clinical significance
- although the available screening tests for classic galactosaemia appeared to be reliable, a number of babies would develop symptoms before the screening process was completed, thereby limiting the impact of screening in reducing the risk of disease complications
- a galactose-restricted diet had not been shown to significantly improve long-term outcomes; moreover there was a lack of evidence that early treatment, as a result of screening, improved outcomes compared with treatment following clinical detection

Findings and gaps in the evidence of this review

Within the scope of this UK NSC evidence summary, 9 unique studies were identified with 1 study providing information relevant to 2 key questions. The overall quantity, quality and direction of the findings was insufficient to answer 2 out of 3 key questions of this evidence review. A summary of question-level results is presented below.

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."

This evidence summary update identified 5 case series providing information on the average age of presentation in classic galactosaemia, 3 from countries where newborn screening is not performed (UK,[7] Croatia[8] and Turkey[9]), and 2 Italian studies[10, 11] evaluating cases identified through from newborn screening programmes.

The studies represented different settings, and the quality varied, with some lacking clarity on the diagnostic criteria used or the comprehensive of representation. For example, the UK study included only children with classic galactosaemia who had presented with acute liver failure to a single UK hospital. Nevertheless, the findings were overall consistent and

suggest that most infants will present clinically by around 7 days of life. Signs of liver dysfunction/ liver failure appear to be universal as the presenting symptom, with sepsis, cataracts and lethargy also frequently reported. The 2 Italian newborn screening studies reported that the majority of cases (>75%) presented clinically before screening results were received or a diagnosis made.

The findings suggest the timing of development from latent to declared disease in classic galactosaemia is reasonably understood, and therefore criterion 1 is met. However, this may raise questions over the benefits of screening given that it may not be able to prevent neonatal toxicity. It is expected that a large proportion of screen-detected cases would be symptomatic by the time screen results are confirmed and diagnosis is made.

Criterion 4 — "There should be a simple, safe, precise and validated screening test."
Criterion 5 — "The distribution of test values in the target population should be known and a suitable cut-off level defined and agreed."

Two retrospective, national newborn screening programme evaluations were identified, from Italy[10] and The Netherlands.[12] These 2 large studies were overall of good quality, with the main uncertainty being the applicability of the index tests and cut-offs to the UK.

In the Italian programme[10] screening involved an initial and repeat total galactose measurement; in the Dutch programme[12] the index tests (total galactose with/without GALT enzyme activity) and cut-offs were continually modified to try and optimise test performance. Both studies demonstrated that all tests had very poor positive predictive value (PPV) for classic galactosaemia, ranging from 0.14% in the Italian programme[10] to a maximum of 6.9% in the Dutch programme.[12] Therefore the majority of screen positives in these screening programmes were false positives. The Italian programme[10] demonstrated that a quarter of screen positives had Duarte galactosaemia, which does not require management or monitoring (the Dutch programme[12] did not detail any diagnoses among false positives). Despite the low PPV (which will be influenced by low population prevalence), the Dutch programme[12] demonstrated that the tests had 100% sensitivity for identifying newborns with classic galactosaemia, with no clinically-diagnosed cases among false negatives. Specificity was also very high with a low false positive rate ranging from 0.49% to 0.02% depending on the test and method used (the Italian programme[10] did not have data available on screen negatives to calculate sensitivity or specificity). These findings of maximum sensitivity and high specificity but generally poor PPV, with the potential for identification of Duarte galactosaemia and other variants of unknown clinical significance, are overall compatible with previous screening programme evaluations. On this basis, there was insufficient evidence to establish an optimal screening test and cut-off to use for classic galactosaemia screening. Therefore, criteria 4 and 5 were not met.

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."

This evidence summary update found a limited body of low quality evidence to inform whether early initiation of treatment improves short- and long-term outcomes in classic galactosaemia. No randomised controlled trials have been conducted that have compared outcomes for screened compared with non-screened individuals (where screening may be a marker for earlier initiation of treatment), as confirmed by one 2017 Cochrane systematic review.[13] Likewise no prospective comparative cohorts have been conducted that have compared outcomes for children by screening status or time of treatment initiation.

Only 2 case series[14, 15] provided evidence for this question. These were the international GalNet registry,[15] which since 2014 has recorded data on 509 patients with classic galactosaemia from 15 countries, and a Turkish case series[14] of 46 patients homozygous for the common pathogenic variant p.Gln188Arg. Both studies were assessed to be low-quality as comparative studies, with various uncertainties and gaps in the data. For example, both studies had the potential for selection bias (such as only including those who had complete assessments documented), the analysis in the GalNet registry lacked absolute numbers and had the potential for confounding (for example, by genotype, country of origin, care provision), and the Turkish study was limited by small sample size and uncertain applicability to the UK. The 2 studies had inconsistent findings. The GalNet registry[15] found that newborn screening and galactose-restriction in the first week of life were both associated with a 70% reduced relative risk of short-term neonatal complications. Newborn screening was also associated with a 70% reduced risk of longer-term neurological outcomes (long-term outcomes were not assessed for early treatment). The Turkish case series[14] conversely found no association between age at diagnosis and risk of longer-term childhood developmental delay or neurological problems.

Overall, the findings of this UK NSC evidence summary are in agreement with the 2015 UK NSC evidence summary, finding a limited quantity of low-quality evidence that has given mixed findings on the association between early treatment and long-term outcomes in classic galactosaemia. Therefore criterion 9 was not met.

Recommendations on screening

Based on the overall synthesis of evidence against the UK NSC criteria, the evidence remains insufficient in volume, quality and consistency to reconsider the current recommendation of not screening for classic galactosaemia.

Evidence uncertainties

Only 2 studies on test accuracy were identified, both of which were evaluations of newborn screening programmes that have been nationally implemented. These used different index tests and cut-off values. In one of the 2 studies, measures of sensitivity or specificity were not reported and could not be calculated. Though further evidence on the sensitivity and specificity of tests to detect classic galactosaemia would be ideal, it is acknowledged that this is difficult in the study of rare diseases. Finding ways to address this is important, particularly given the potential for identification of Duarte galactosaemia and other variants of unknown clinical significance. Further screening studies with improved methodological consistency (in terms of index test cut-offs, repeat testing and the reference standard) may be achievable and would allow for an informative evaluation of a test to be used in newborn screening for classic galactosaemia.

In addition, there was limited low quality evidence from 2 retrospective studies to assess whether early treatment improves short- and long-term outcomes in classic galactosaemia. Neither of these studies was designed as a comparative study to address this question, and both had quality and applicability limitations. Consistency in study methodology and investigated treatment outcomes in future research may mitigate some of the current gaps and uncertainties presented within the current evidence base. Prospective cohort studies comparing screening with no screening could be valuable. This could help to assess whether screening enables earlier detection and treatment initiation prior to symptomatic presentation and whether it reduces the risk of short-term and long-term outcomes.

Limitations

This UK NSC evidence summary was conducted using a rapid review methodology. The searches were limited to 3 literature databases and did not include grey literature resources. This evidence review included only peer-reviewed journal publications in the English language. The reviewers were also unable to contact study authors or review non-published material. However, this is an accepted methodological adjustment for a rapid review, and these limitations should not have led to the exclusion of any pivotal studies.

Introduction and approach

Background

Galactosaemia is a hereditary disorder of galactose metabolism, a type of condition known as an inborn error of metabolism (IEM).

Aetiology and epidemiology

The predominant source of galactose is lactose in milk, which is made up of glucose in combination with galactose. A small amount of galactose is also made endogenously within the body. Galactose is broken down in the Leloir metabolic pathway[2] which involves 3 enzymes. The first enzyme in the pathway is galactokinase (GALK), which converts galactose into galactose-1-phosphate (Gal-1-P). The second enzyme galactose-1-phosphate uridyltransferase (GALT) then converts Gal-1-P and uridine diphosphate (UDP)-glucose into UDP-galactose. The final enzyme in the pathway, uridine diphosphate galactose-4-epimerase (GALE), converts UDP-galactose back into UDP-glucose, which is then fed back into the metabolic pathway.[2] Galactosaemia can be caused by deficiencies in any one of these enzymes, GALK, GALT or GALE. The condition being reviewed here, and the potential candidate for newborn screening, is classic galactosaemia, caused by deficiency of the second enzyme, GALT. This results in accumulation of galactose and its metabolites in the blood, which have toxic effects. Deficiencies of the other 2 enzymes, GALK and GALE, cause conditions that are biochemically and clinically distinct from classic galactosaemia and thus fall outside of the scope of this evidence review.[2]

Classic galactosaemia is caused by abnormal variants (mutations) of the *GALT* gene, which is positioned on chromosome 9p13.[2, 5] The condition has an autosomal recessive pattern of inheritance, meaning that individuals with classic galactosaemia have inherited 2 disease-causing (pathogenic) variants, one from each parent. A total 340 *GALT* variants have been identified to date,[16] many of which are established to be pathogenic (including deletions, missense, splice site and frameshift variants), while others are understood to be silent/benign or associated with mild symptoms only.

The incidence of classic galactosaemia will be affected by the frequency of pathogenic variants in the population, and factors such as consanguinity. A UK screening pilot from 1988 to 1990 reported the UK incidence to be 1 in 44,000 newborns.[3] No more recent UK incidence estimates are available. However, a 2013 Irish study reported an incidence of around 3 in 100,000 (or 1 in 34,000) newborns among the non-Traveller population but 232 per 100,000 (or 1 in 430) among the Traveller community.[4] Published studies from other European countries (Germany, Spain, Denmark, Hungary, Austria and Greece) have given

incidence estimates ranging from 1 per 100,000 (Sweden and Denmark) to 4.5 per 100,000 newborns (Greece).[4] In the US, the incidence of classic galactosaemia has been reported at around 1.5 to 2 per 100,000 newborns.[2, 4]

The international clinical guideline of classic galactosaemia states that the condition can be definitively diagnosed by: [5]

- absent or profoundly deficient GALT enzyme activity in red blood cells
- and/or *GALT* gene analysis demonstrating 2 pathogenic variants (which is sufficient for the diagnosis alone)

The guideline applies the term classic galactosaemia to those with GALT level of less than 15%.[5] However, this is largely on account of the wider spectrum of compound heterozygotes carrying milder variants who may be detected through newborn screening programmes. Most individuals with disease complications will have 'absent or profoundly deficient' GALT enzyme levels of less than 1%.[17]

As of July 2018, the International Galactosaemia Network registry[15] (GalNet), established in 2014, had collected data for 509 individuals with classic galactosaemia from 15 countries (13 European, including UK, US and Israel). This registry has included only those with GALT activity ≤10% (and/or pathogenic variants). The majority of all individuals have GALT enzyme activity of ≤1% (83%), 14% have GALT activity 1 to 5%, and only 3% have enzyme activity of 5 to 10%.[15] The most common genotype among this 94% Caucasian/European population is homozygosity for the pathogenic variant c.563A>G (p.Gln188Arg) (58%).[15]

Clinical presentation and relationship to genotype

Classic galactosaemia with profoundly deficient GALT activity usually presents in the first weeks of life with signs of liver dysfunction such as jaundice and hepatomegaly, lethargy, poor feeding, vomiting, diarrhoea and sepsis.[2] The various systemic effects are understood to be caused by excess accumulation of galactose and its metabolites Gal-1-Pand galactitol.[17] Eighty percent of cases in the GalNet registry had developed neonatal illness, with the most common presenting signs/symptoms being:[15]

- elevated liver enzymes (70%)
- bleeding diathesis/abnormal clotting (43%)
- encephalopathy (29%)
- infection (27%)
- cataract (26%)
- hypoglycaemia (25%)

Compound heterozygotes who carry one pathogenic gene variant, such as c.563A>G (p.Gln188Arg), but another variant that confers GALT enzyme activity greater than 1% are

understood to have less severe neonatal disease and better long-term prognosis.[17] The international guideline highlights c.404C>T (p.S135L) as one such variant (more common in those of African American descent[17]) which demonstrates enzyme activity of only 1 to 2% in red blood cells but up to 10% in the liver and other tissues.[5]

A condition termed Duarte galactosaemia or partial galactosaemia occurs in those who carry one pathogenic variant and one Duarte variant (of which 5 have been identified to date).[5, 16] Individuals with Duarte galactosaemia demonstrate GALT enzyme activity of 14 to 25% and are not known to develop clinical symptoms.[5] However, these individuals are detected through newborn screening programmes and there has been debate on the appropriate management. European countries are reported to perform no follow-up of individuals with Duarte, but some US centres are reported to prescribe galactose-restriction.[5] The international clinical guideline recommends that doctors treat individuals with classic galactosaemia who have a red blood cell GALT activity level <10% and/or carry 2 pathogenic variants (specified to include p.S135L). However, the guideline reports that there is insufficient evidence to support the treatment of individuals with enzyme activity level of 10 to 15%. It also specifically advises against treating Duarte galactosaemia.[5]

Management and prognosis

For classic galactosaemia (GALT activity below 10 to 15%), management involves a galactose-restricted diet. As lactose is the predominant dietary source, breastfeeding and whey-based formula need to be discontinued upon suspicion of the diagnosis, and replaced with soy-based or elemental formula.[5] After weaning, all animal milks (including 'lactose-free') and milk products need to be avoided, and this is continued for life. However, there has been practice variability in permission of foods that contain trace levels.[5, 18] In 2012, the Task Force of the Galactosemia Foundation (US) reviewed the available literature on the galactose content of different foods.[19] The vast majority of fruit and vegetables, legumes, non-fermented soy products (that is, excluding soy sauce, miso and fermented tofu), and aged hard cheeses (including Parmesan and sharp Cheddar) contain less than 50mg galactose per 100g and may be permitted.[19] This is in contrast to milk which contains around 2400mg galactose per 100 ml.[19] These trace levels are also minor compared with the endogenous synthesis of galactose in the body, ranging from over 25mg/kg/day in babies decreasing to 8mg/kg/day in adults.[5]

Even with strict galactose-restriction, classic galactosaemia is associated with long-term complications including neurodevelopmental problems, poor bone mineral density and ovarian failure in females. The mechanisms behind this are not fully understood, though it is thought that glycosylation may not normalise in many patients, despite galactoserestriction.[17] All patients in the GalNet registry received soy-based formula (or alternatives) in infancy, which was implemented within the first week of life for 50% and

within 2 weeks for 85% of patients. Nearly all (94%) have continued to follow a lactose-free diet throughout life (most with no restriction of non-dairy sources).[15] Despite this, 50% of patients are recorded to have global developmental delay, and two-thirds language and speech disorders. Neurological complications are reported to affect half, with tremor and motor problems being the most common, and just under half had psychological and behavioural problems. Half of female patients had experienced delayed or induced puberty.[15]

The international clinical guideline gives extensive recommendations for ongoing monitoring and follow-up of individuals diagnosed with classic galactosaemia. They advise that red blood cell levels of Gal-1-P are assessed at 3 and 9 months after initiation of galactose-restriction, and then annually until an individual baseline level is established.[5] Numerous recommendations are then given on tests of neurodevelopment and intellect throughout childhood, with endocrinology follow-up, assessment of bone health and other potential systemic complications.[5]

Screening for classic galactosaemia

Classic galactosaemia requires a galactose-restricted diet and this needs to be initiated as soon as possible to try and reduce the risk of complications. A 2016 systematic review (Varela-Lema et al)[4] reported that an international survey of 371 patients from Europe and the US had found that 91% of infants who were treated from birth because of an affected sibling did not develop newborn symptoms.[4] However, universal pre-emptive galactose-restriction from birth would clearly not be feasible or ethical. The value of newborn screening may therefore be debated given that galactose exposure in the first week of life may result in neonatal toxicity before the screening process has been completed. Past research into screening for galactosaemia has given mixed findings, though the evidence base on the whole appears to be quite limited. Both the Varela-Lema et al systematic review[4] and Seymour et al health technology appraisal (HTA),[3] published 20 years earlier, report the only UK screening study to be the British Paediatric Surveillance Unit Study, conducted between 1988 and 1990 (Honeyman et al[20]). This study found no difference in the proportion of screened and non-screened infants who presented with severe symptoms, and no indication that screening reduced neonatal mortality. On this basis, Honeyman et al had concluded that there was no justification to screen for classic galactosaemia in the UK.[3] Since then, the Varela-Lema et al (2016)[4] report that screening programmes from Sweden, Italy and the south-west of Germany have observed reduced neonatal mortality from galactosaemia since the introduction of screening compared with previously. However, in contrast, Varela-Lema et al[4] reference 6 studies (published pre-2014) that observed no association between the incidence of long-term complications and the age of diagnosis, time of initiation of dietary management, or intensity of galactose restriction.[4]

Current global landscape of newborn screening for classic galactosaemia

Reflecting this variable evidence, there is currently wide international variation in screening for classic galactosaemia. A 2016 review of newborn screening for inborn errors of metabolism[6] reported that classic galactosaemia screening was performed in all US states, 8 of 15 Canadian territories, 14 of 20 countries in Latin America, 9 of 24 countries in Asia Pacific, and 13 of 48 European countries (not specified). The Varela-Lema et al systematic review[4] (search date 2014) reported that publications on screening programmes for galactosaemia were available from 8 European countries (in addition to the US and Japan):

- Ireland
- Spain
- Germany (publications for 2 regions)
- Austria
- Hungary
- Greece
- Denmark
- Sweden

A narrative review from 2017[21] reported that Denmark has actually discontinued screening, with Scotland and Norway also having discontinued programmes. Switzerland is mentioned by this narrative as being another country that has an established programme, while Belgium, Italy and Turkey are reported to have pilot programmes.[21] However, it is unclear whether this is the latest comprehensive list on the European screening landscape.

The screening process and prior evidence on test performance

As a process, newborn screening for classic galactosaemia is relatively safe and simple to perform: total galactose (galactose plus Gal-1-P) or GALT enzyme activity can be measured in the newborn dried blood spot (DBS) sample.[3, 4] Various laboratory assays have reported use, including tandem mass spectrometry, fluorometric and colorimetric assay and newer ultra-performance liquid chromatography methods. The European screening programmes currently established vary considerably in their specific methods. The 9 European publications identified by Varela-Lema et al[4] showed that these programmes screened at between 1 and 9 days of life, used various assays and cut-offs for total galactose or GALT measurement, and variable methods of diagnostic confirmation. Some programmes used a single screening approach while others used a several-tiered screening approach combining several measurements and/or assays. The ‘age at diagnosis’ was variably reported across studies: Austria reported around 7 days, Sweden ≤7 days, Spain 10 days, Greece 7 to 15 days and one of the German publications, 8 days. Denmark specified 2 to 7 days after samples, while the other German publication showed that 57% diagnosed within 48 hours of sampling.[4]

Five of the screening programme evaluations identified by Varela-Lema et al[4] (4 dated 2011/2012 and one dated 2002) had provided test performance data or allowed its calculation. Reported test performance ranged from a false positive rate (FPR) of 0.25% and positive predictive value (PPV) of 0.9% (for the Austrian programme using quantitative measurement of total galactose), to a FPR of 0.0005% and PPV of 64% (in the 2-tiered Swedish programme combining total galactose and GALT measurement).[4] The Seymour et al HTA (1997)[3] had reported the findings from an earlier Dutch programme, which found an FPR of 0.028% (by colorimetric assay measurement of total galactose). Forty percent of whom were compound heterozygotes, mostly with partial/Duarte galactosaemia.[3] As reported, the international clinical guideline does not currently recommend management for these individuals. None of these screening programme evaluations reported any false negatives (all with estimated sensitivity 100%). However, the HTA[3] had cautioned there may be the possibility of false negatives when storing and transporting dried blood spot (DBS) samples for GALT measurement, as enzyme activity can degrade under certain temperature and humidity conditions.

Previous conclusions on the benefits and drawbacks of screening

Overall, the 1997 HTA[3] had concluded that classic galactosaemia was a well-defined disorder with known UK incidence and one that was associated with significant morbidity. However, it pointed out that a screening programme could not be supported in the UK given that there were outstanding questions around whether there was a safe, simple and robust screening test and uncertainty whether earlier treatment could improve outcomes.

The Varela Lema et al systematic review[4] similarly concluded the severity of classic galactosaemia, and that treatment during the asymptomatic phase may prevent the most severe complications and deaths. However, the authors considered that in order to be successful, the screening process would need to be completed and treatment commenced within the first week of life. Nevertheless, in line with the findings of the HTA, they noted that studies have consistently demonstrated patients to suffer from long-term disability despite appropriate treatment. Another concern highlighted by this systematic review was the variability in tests and cut-offs used in current screening programmes, and the potential harms from false positives, including unnecessary dietary restriction for those with Duarte galactosaemia.

Of the IEMs, there is more or less international consensus only in screening for phenylketonuria (PKU).[4, 6] This is a condition where high levels of phenylalanine causes irreversible brain damage, and early treatment is known to prevent neurodevelopmental impairments.[6] A range of other IEMs are variably included in newborn screening programmes worldwide. The NHS newborn blood spot screening programme currently

includes 5 others alongside PKU: medium-chain acyl-CoA dehydrogenase deficiency (MCADD), maple syrup urine disease (MSUD), isovaleric acidaemia (IVA), glutaric aciduria type 1 (GA1) and homocystinuria (HCU).

Current policy context and previous reviews

The UK NSC does not currently recommend newborn screening for classic galactosaemia. This recommendation was made on the basis of the last UK NS evidence summary on the topic, published in 2015.

The last UK NSC evidence summary had explored the volume, quality and direction of the literature published since the 1997 HTA[3] and concluded that:

- while classic galactosaemia was sufficiently understood, a screening programme may identify individuals carrying variants of uncertain clinical significance
- although the available screening tests for classic galactosaemia appeared to be reliable, a number of babies would develop symptoms before the screening process was completed, so limiting the impact of screening in reducing the risk of disease complications
- a galactose-restricted diet had not been shown to significantly improve long-term outcomes; moreover there was a lack of evidence that early treatment as a result of screening improved outcomes compared with treatment following clinical detection

Objectives

The current UK National Screening Committee (NSC) evidence summary aims to assess whether there has been new evidence, published since the 2015 UK NSC evidence summary,[1] on the age of clinical presentation, the accuracy of available screening tests for classic galactosaemia, and whether early treatment improves short-term and long-term outcomes.

Three key questions will be addressed to cover the issues identified by the last UK NSC evidence summary. These questions are outlined in Table 1 alongside the criteria set out by the UK NSC for assessing the suitability of a screening programme.

Table 1. Key questions for the evidence summary, and relationship to UK NSC screening criteria

Criterion	Key questions	Studies Included
THE CONDITION		

	Criterion	Key questions	Studies Included
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.	1. What is the median age of presentation of classic galactosaemia?	5 retrospective case series (2 within the setting of NBS programmes for galactosaemia)
	THE TEST		
4 **5**	There should be a simple, safe, precise and validated screening test. The distribution of test values in the target population should be known and a suitable cut-off level defined and agreed.	2. What is the accuracy of the available screening tests to detect classic galactosaemia?	2 retrospective case series (NBS programme evaluations)
	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.	3. Does early initiation of treatment for individuals with classic galactosaemia provide better short- and long-term outcomes?	1 systematic review of RCTs, 1 registry study and 1 case series

Abbreviations: RCTs – randomised controlled trials

Methods

The current review was conducted by Bazian (part of the Economist Intelligence Healthcare Unit), in keeping with the UK NSC evidence review process.

Databases/sources searched

A systematic search of literature search of MEDLINE and Embase databases (via Embase.com) and The Cochrane Library (via Wiley Online) was conducted on 22nd January 2020 to identify studies relevant to the questions detailed in Table 1 that had been published since January 2014 (the search date for the last UK NSC evidence summary). Due to the small body of evidence, the search was additionally run in Pubmed on 18th February as a check to ensure that this did not contain any citations unlisted in Embase (no additional relevant literature was identified).

Eligibility for inclusion in the review

The *a priori* inclusion/exclusion criteria for each key question are outlined in Table 2. Searches were conducted using high-level index terms related to the disease, rather than conducting a focused search around the PICO (population, intervention, comparator, outcome) for each question due to the expected small volume of evidence. Searches retrieved a total of 395 citations after removal of 140 duplicates. The full search strategy is presented in Appendix 1 (Tables 9 to 12).

The following review process was followed:

1. Each of the 395 titles and abstracts were reviewed against the inclusion/exclusion criteria for each question by one information specialist. Where the applicability was unclear the article was included at this stage to ensure that all potentially relevant studies were captured. Studies providing potential background/contextual information were also included. In total, 69 citations were included at first sift.
2. At second sift the main reviewer reviewed each of the 69 abstracts for potential relevance to any of the 3 questions. Where the article content was unclear from the abstract, full text was obtained to ensure that potentially relevant literature was not missed.
3. A total of 22 articles were acquired for the full-text review stage. Each full-text article was reviewed against the inclusion/exclusion criteria by the main reviewer, who determined whether the article was relevant to any of the 3 review questions, or provided background context. All inclusion/exclusion decisions were reviewed by a second independent reviewer who provided input in cases of uncertainty. Any disagreements were resolved by discussion until a consensus was reached.
4. The citations of the retrieved studies were additionally hand-searched to check that no relevant studies had been missed.

Table 2 outlines the exclusions by study design for each key question (for example, conference abstracts). The following exclusions were then applied on appraisal of the content of the retrieved literature on classic galactosaemia, either at abstract level, or at full text review if the content was unclear from the abstract. These broad exclusions have been listed here, rather than in the individual criteria sections, as most are not specific to question. This included studies:

- conducted exclusively in Middle Eastern, Asian or Latin populations with uncertain applicability to the UK
- presenting/validating new analysers to screen for inborn errors of metabolism
- validating potential new methods to monitor treatment response
- developing methods to measure genotype from dried blood samples (DBS) as oppose to need for whole blood samples
- looking at degradation of enzyme activity in DBS under different storage conditions (for example, by time and temperature)
- analysing the association between phenotype and enzyme activity, galactose index or genotype (for example, understanding the significance of unique variants)
- analysing the frequency of galactosaemia gene variant carriage among a general population sample (one study identified from Greece)
- assessing endogenous galactose production
- analysing different food samples for galactose content
- analysing care needs in galactosaemia (using the Capacity Profile)
- assessing the maternal impact of replacement of breastfeeding
- individual case reports (including those also analysing parents/siblings)
- exclusively analysing Duarte (partial) galactosaemia (for example, comparing outcomes between cases and healthy controls)

0contains a full PRISMA flow diagram (Figure 2), along with summary tables of the studies reviewed at full-text. This includes a table of publications included for the 3 key questions (Table 13), and a table of publications that did not provide evidence applicable to the key questions, with rationale for exclusion (Table 14).

The reviewers had also planned *a priori* to identify any background information on the number of cases of classic galactosaemia identified as part of the current UK newborn screening programme for phenylketonuria (PKU). No information on this was identified through the retrieved literature. Contact with the Public Health England screening data and management team confirmed that such information is not routinely documented.

Table 2. Inclusion and exclusion criteria for the key questions

Key question	Inclusion criteria							Exclusion criteria
	Population	Target condition	Intervention	Reference Standard	Comparator	Outcome	Study type	
1. What is the median age of presentation of classic galactosaemia?	Newborns from western populations applicable to the UK	Classic galactosaemia	NA	NA	NA	Median (or mean) age at presentation	Cross sectional studies or cohort studies of the newborn population Cohorts or case series of children with classic galactosaemia	Individual case reports, conference reports, abstracts, editorials, non-English language studies.
2. What is the accuracy of the available screening tests to detect classic galactosaemia?	Newborns from western populations applicable to the UK	Classic galactosaemia	Any screening markers as used by applicable studies: Total galactose or Gal-1-P levels GALT activity level Genetic testing	Gold standard as used by individual studies or genetic testing	None or any	Sensitivity, specificity, PPV, NPV, positive LR, negative LR	Diagnostic cohort studies of randomly-selected or consecutively enrolled populations reporting test performance data	Case-control studies (though any identified would be reviewed in the absence of applicable studies), conference reports, abstracts, editorials, non-English language studies.

3. Does early initiation of treatment for individuals with classic galactosaemia provide better short- and long-term outcomes?	Newborns from western populations applicable to the UK	Classic galactosaemia	Early initiation of galactose-restricted diet	NA	Later initiation of diet (as defined by the individual study)	Any health outcomes including neuro-developmental problems, liver disease, cataracts, mortality	Randomised controlled trials or comparative cohort studies	Non-comparative cohort studies and case series (though any identified would be reviewed in the absence of applicable studies), conference reports, abstracts, editorials, non-English language studies.

Abbreviations: Gal-1-P, galactose-1-phosphate; GALT, galactose-1-phosphate uridyltransferase; PPV, positive predictive value; NPV, negative predictive value; LR, likelihood ratio; NA, not applicable

Appraisal for quality/risk of bias tool

The majority of included studies were cases series or retrospective non-comparative screening programme evaluations. Five of these case series[7-11] provided evidence for question 1 on age at presentation. They were assessed using a quality assessment tool for cases series developed by Murad et al[22] and based on adaption of the Newcastle Ottawa scale for cohort and case-control studies (with removal of items related to comparability and adjustment), Bradford Hills and Pierson criteria. Of note the overall judgement on quality and applicability was primarily based on relation of for the evidence for the key question and does not necessarily represent the quality of the study for its overall intended purpose.

One of these case series/screening programme evaluations[10] additionally provided data which allowed calculation of some aspects of screening test performance. Therefore this study was additionally appraised using the QUADAS-2 tool for assessment of diagnostic accuracy studies. A further programme evaluation[12] had primarily aimed to review the effectiveness of its screening methods, and contained no information applicable to question 1 or 3. This study was therefore appraised using only the QUADAS-2 tool.

Three studies provided evidence for question 3. One was a systematic review[13] which was appraised using the Critical Appraisal Skills Programme (CASP) Systematic Review Checklist. The second was a retrospective review of patient information collected to date in the previously mentioned GalNet Registry[15] and the third a case series of patients with a specific genotype seen at a Turkish hospital.[14] Both of these studies were primarily non-controlled cohort studies. However, they did contain some evidence applicable to the question of whether screening or early treatment improves outcomes compared with no screening or late treatment. The quality of the evidence for this purpose was therefore assessed using the CASP Cohort Checklist for comparative studies. However, this assessment does not diminish from the value of these studies for their intended purpose as a retrospective evaluation of, respectively, the characteristics of cases within the GalNet registry[15] and those seen at a single centre.[14]

The individual quality assessments are presented in the ‘Summary and appraisal of individual studies’, Appendix 3, Tables 18 to 27.

Question level synthesis

Criterion 1 — Median age of presentation of classic galactosaemia

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."

Question 1 – What is the median age of presentation of classic galactosaemia?

Understanding the typical age of symptomatic presentation in classic galactosaemia is important because it helps to determine the potential impact of a newborn screening (NBS) for classic galactosaemia programme. In the UK, the newborn blood spot is currently taken on day 5 which will allow galactose exposure in the first days or even weeks of life until the screening results are received. Even if treatment was started immediately upon disease suspicion (for example, upon screen-positive results rather than waiting until diagnostic confirmation) this may still allow for galactose toxicity and newborn complications to occur. If newborn symptoms typically develop before the anticipated time of screening or diagnosis, then the NBS screening programme may have limited effectiveness in preventing disease complications. On the other hand, if symptoms would normally develop after the time of potential screening, then screening may be able to have greater impact in reducing disease complications.

The last 2015 UK NSC evidence summary[1] had found that it was not possible to identify a median age of symptomatic presentation. The evidence summary identified 3 studies of newborn screening programmes in Europe, where NBS was performed at 3 to 5 days of life. Between 22% and 75% of newborns across these studies were symptomatic at the time of diagnosis. A single study was identified outside of the newborn screening context which reported that 47% of cases presented within the first 2 weeks of life. The identified studies did not specify the nature or severity of symptoms. Overall, the 2015 UK NSC evidence summary concluded that the body of evidence was limited, but suggested that a substantial proportion of screen-detected cases will be symptomatic by the time screen results are confirmed and diagnosis is made.

This UK NSC evidence summary aimed to see whether there was new evidence on the median age at symptomatic presentation of classic galactosaemia which may help to inform the potential benefit of a screening programme.

Eligibility for inclusion in the review

This UK NSC evidence summary aimed to identify cohort studies or case series of newborns with classic galactosaemia that contained information on the age at or timing of symptomatic presentation. Cases could either be clinically-detected or identified through newborn screening programmes. Studies could be from the UK or of representative western/European populations. Studies conducted exclusively in Middle Eastern, Asian, African or Latin populations were excluded; although a single applicable case series from Turkey was included on the basis of this being a borderline European/Middle Eastern country (a Turkish study had also been included in the last UK NSC evidence summary). The decision was made *a priori* that any case reports identified by the literature search would be reviewed on an individual basis for potential inclusion. However, the only case reports retrieved came from non-western populations so were excluded.

Systematic reviews on classic galactosaemia were reviewed to see if they contained information on age at or timing presentation. However, if any such information was only cited to studies published prior to the 2014 search date, the systematic review was excluded as it was not considered to be providing new evidence. Other exclusions for reasons of being out of question scope have been listed in the methods section.

Description of the evidence

Of the 22 publications appraised at full text across for this evidence summary, 5 case series were judged to contain evidence applicable to this question. Two studies[10, 11] were retrospective reviews of cases of classic galactosaemia identified through NBS programmes and two studies[8, 9] were case series of patients with classic galactosaemia who were clinically detected. The only UK study[7] was a case series of infants presenting to a London hospital with acute liver failure, some of whom had classic galactosaemia. The sample size of cases with classic galactosaemia ranged from 13 to 76 across studies.

The findings from these studies are summarised in Table 3, with full evidence extraction in Appendix 3, Table 15. Quality assessment is summarised in Table 4, with the detailed appraisal for each study given in Appendix 3, Tables 18 to 22. The reviewers used the appraisal tool for case series as suggested by Murad et al,[22] which is based on the Newcastle Ottawa scale for cohort and case-control studies (with removal of items related to comparability and adjustment), Bradford Hills and Pierson criteria. Notably, assessments were based on the quality and applicability of the study content to address the key question; they do not represent the quality of the study for its intended purpose as a descriptive case series.

Table 3. Case series providing data on age at clinical presentation in classic galactosaemia

Study, country	Setting/context	Sample	Method of diagnostic confirmation	Age at presentation	Other measure on timing of presentation	Presenting symptoms
Hegarty et al 2015[7] **Single centre, UK**	Non-screening Children aged <5 presenting with acute liver failure 2001 to 2011	n=17	GALT level (values/cut-off NR) Gene analysis NR	Median 7 days	NR	100% jaundice 94% hepatomegaly 18% hepatic encephalopathy 13% splenomegaly
Porta et al 2015[10] **Regional centre, Italy**	NBS mean 3.4 days recall 12.2 days (treatment start) Case review 1982 to 2012	n=13	GALT level (mean 2.1%, cut-off NR) Following 2x TGal <10mg/dL (initial NBS + recall) Gene analysis NR	Mean 5.8 (+/- 1) days	77% (n=10) readmitted to hospital before NBS results	Described as uniform: jaundice, lethargy, vomiting and liver failure
Ramadza et al 2018[8] **Croatia (unclear setting)**	Non-screening Case review time period NR (though cases up to 40 years of age)	n=16	Reduced GALT and raised TGal (values/cut-off NR) Genotypes given	Range 2 days to 6 weeks: 50% ≤7 days 19% 8 to 14 days 13% 15 to 21 days 6% 6 weeks 6% unspecified 6% pre-emptive treatment	NR	Common across all: raised bilirubin, coagulopathy, other liver dysfunction and sepsis. Additional reports of lethargy and cataracts
Teke Kisa et al 2019[9] **4 centres, Turkey**	Non-screening Case review 1996 to 2017	n=76	GALT level <10%; plus 2 pathogenic variants	Median 10 days (25^{th} to 75^{th} centile 5 to 20 days)	NR	'At presentation or time of diagnosis': 92% jaundice 66% hepatomegaly 54% coagulopathy 43% sepsis 34% cataract
Viggiano et al 2015[11]	NBS	n=14	GALT level <5%; plus 2	NR	93% (n=13) reported to be 'symptomatic at	100% liver failure 62% cataracts 38% sepsis

Regional centre, Italy	time of screen, diagnosis or treatment NR Case review 1980 to date		pathogenic variants		the time of diagnosis'	31% anaemia

Abbreviations: GALT, galactose-1-phosphate uridyltransferase; NBS, newborn screening; NR, not reported; TGal, total galactose

Table 4. Quality assessment of case series (criteria described by Murad et al[22])

Domain	Hegarty et al[7]	Porta et al[10]	Ramadza et al[8]	Teke Kisa et al[9]	Viggiano et al[11]
Selection[1]	Low	Low	Unclear	Low	Low
Ascertainment[2]	Unclear	Low	Low	Low	High
Causality[3]	NA	NA	NA	NA	NA
Reporting[4]	High	Low	Unclear	Unclear	Unclear
Overall assessment of quality and applicability for KQ1 (with summary rationale for downgrading)	Fair/poor lack of reporting of diagnostic criteria and narrow representation of only cases who present with liver failure	Good	Fair clear data on diagnostic criteria and age at presentation, but some uncertainties around diagnostic criteria and representation	Good/Fair good quality – except for high consanguinity and incidence among this Turkish population	Fair/poor diagnostic criteria are reliable but lacking detail on specific age at presentation

[1] Selection: overall representation of participants
[2] Ascertainment: assessment of exposure (here considered as diagnosis of classic galactosaemia) and outcome (considered as presenting age)
[3] Causality: domain mostly applicable to reporting of drug adverse events therefore not applicable (NA) to this assessment
[4] Reporting: sufficient detail to allow replication of research, and practitioners to make inferences related to practice

Discussion of findings

The identified studies present a reasonable volume of new evidence on the age of clinical presentation in classic galactosaemia. The studies represent a heterogeneous group of settings and differ in the methods of data collection, but overall the findings seem consistent. Clinical presentation is reported at a mean or median of 6, 7 and 10 days in 3 studies. One study reported a wide range of 2 days to 6 weeks, but 50% presented at up to 7 days and 69% by 14 days. Therefore the evidence suggests that as a very rough average most infants will present clinically by around 7 days of age. Across all studies, signs of liver dysfunction/ liver failure appear to be universal as the presenting symptom, with sepsis, cataracts and lethargy also frequently reported.

In terms of the quality and applicability of the evidence, 2 regional studies from Italy[10, 11] represent the newborn screening context. Porta et al[10] was a good quality study, giving clear description of the screening process, diagnostic confirmation and age at presentation (mean 5.8 days). NBS was performed at a mean 3.4 days. However, galactose-restriction was not implemented for screen-positives until they had a repeat confirmatory test, at a mean 12.2 days. Three-quarters of all cases (10/13) were said to be re-admitted symptomatically before NBS results were available. The second Italian study (Viggiano et al[11]) was rated as fair/poor quality evidence for this question. It is expected to be representative of all cases in this region and classic galactosaemia was diagnosed according to gold standard criteria (profound GALT deficiency plus 2 pathogenic variants), but the screening method and timing was not described, nor the age at symptomatic presentation given. Nearly all infants were said to be symptomatic at the time of diagnosis. Given that this is the Italian NBS programme, it may be that the screening process was as described by Porta et al[10]: bloodspot at around 3 days with screen-positives recalled for a confirmatory test at 12 days. In this case, these results could infer that most infants were symptomatic by around 7 to 14 days, though it is not possible to be sure that diagnosis did not refer to the time of later GALT measurement and genotyping. Nevertheless, on balance, both of these Italian studies suggest that in the majority of cases, galactose toxicity has occurred before the NBS process is complete. Notably in Italy newborn bloodspot screening appears to be performed slightly earlier than it would be in the UK, at an average of 3 rather than 5 days. This may suggest there may be even higher likelihood of clinical presentation before NBS results in a potential UK programme (though it is not possible to know what a potential screening process in the UK might involve or when treatment would be initiated).

The remaining 3 studies, conducted in countries where galactosaemia screening is not performed, were a mix of quality. The Turkish case series[9] was good quality, covering a large representative sample of all cases presenting in Turkey over a 20 year period, using gold standard diagnostic criteria and with clear data applicable to the question. The only

potential issue is applicability to the UK of this European/Middle Eastern population, which showed high incidence of classic galactosaemia (1 in 23,775), possibly reflecting high consanguinity. The Croatian case series[8] was of fair quality, giving description of each individual case, their genotyping and applicable information on presentation, but with some uncertainties around patient selection and representation.

The final case series was the only applicable study identified from the UK.[7] This study reviewed all young children (n=127) who had presented to King's College London over 10 years with acute liver failure. Just over a quarter of these presentations (n=36) were due to IEMs, of which classic galactosaemia was the most common, accounting for nearly half of IEMs detected (n=17) and 13% of all liver failure presentations. The median age of presentation for infants with classic galactosaemia (median 7 days) was younger than infants with other IMDs (median 6 weeks). It was rated as low quality for the purposes of this question, as the method of diagnosis was not reported, and because this only represented cases who met criteria for acute liver failure at this single centre. Therefore it is unclear how representative this is of classic galactosaemia presentations in this centre or the wider UK, who may not all meet this criteria.

Summary of Findings Relevant to Criterion 1: Criterion met*

This evidence summary update identified 5 applicable case series providing information on the average age of presentation in classic galactosaemia, 3 from countries where NBS is not performed (UK, Croatia and Turkey), and 2 Italian studies from NBS programmes.

The studies represented different settings, and quality varied, with some lacking clarity on the diagnostic criteria used or the comprehensive of representation. For example, the UK study included only children with classic galactosaemia who had presented with acute liver failure to a single UK hospital. Nevertheless the findings were overall consistent and suggest that most infants will present clinically by around 7 days of life. Signs of liver dysfunction/ liver failure appear to be universal as the presenting symptom, with sepsis, cataracts and lethargy also frequently reported. The 2 Italian NBS studies reported that the majority of cases (>75%) presented clinically before NBS results were received or diagnosis made.

The findings suggest the timing of development from latent to declared disease in classic galactosaemia is reasonably well understood, but this may raise questions over the benefits of screening given that it may not be able to prevent neonatal toxicity. It is expected that a large proportion of screen-detected cases would be symptomatic by the time screen results are confirmed and diagnosis is made.

* **Met** -for example, this should be applied in circumstances in which there is a sufficient volume of evidence of sufficient quality to judge an outcome or effect which is unlikely to be changed by further research or systematic review.
Not Met - for example, this should be applied in circumstances where there is insufficient evidence to clearly judge an outcome or effect or where there is sufficient evidence of poor performance.
Uncertain -for example, this should be applied in circumstances in which the constraints of an evidence summary prevent a reliable answer to the question. An example of this may be when the need for a systematic review and meta-analysis is identified by the rapid review.

Criteria 4 and 5 — Accuracy of screening tests

Criterion 4 – "There should be a simple, safe, precise and validated screening test."
Criterion 5 — "The distribution of test values in the target population should be known and a suitable cut-off level defined and agreed."

Question 2 – What is the accuracy of the available screening tests to detect classic galactosaemia?

The 2015 UK NSC evidence summary[1] identified 7 studies which did not contain sufficient data by which to assess the performance of screening tests for classic galactosaemia. The studies had measured galactose, galactose-1-phosphate (Gal-1-P) or GALT enzyme activity (alone or in combination) and were heterogeneous in the assay used, cut-offs and reference standard. Limited information was available on the number of newborns screened, screen-positives and negatives. Where positive predictive values (PPV) could be calculated this varied widely from 4 to 87%. Some studies also reported between 11% and 67% of screen-positives to have partial/Duarte galactosaemia or other variants of unknown clinical significance, depending on the test and cut-off used. The earlier 1997 HTA[3] had similar findings, that although screening tests appeared to have perfect sensitivity, a substantial proportion of compound heterozygotes would also be identified.

This evidence update aimed to review whether new evidence had been published since 2014 on the performance of screening tests for classic galactosaemia.

Eligibility for inclusion in the review

The reviewers aimed to identify diagnostic cohort studies conducted in consecutively enrolled or randomly selected samples of newborns from representative western/European populations. Studies could use any screening marker or assay to test for classic galactosaemia using dried blood spot samples including measurement of total galactose, Gal-1-P, GALT enzyme activity, or DNA testing for *GALT* variants. The reviewers aimed to consider any reference standard as used by eligible publications. However, demonstration of absent or profoundly deficient GALT activity in red blood cells and/or carriage of 2 pathogenic variants would be considered the gold standard, as recommended by the International clinical guideline of classic galactosaemia.[5] Eligible studies were required to provide data on test performance including sensitivity, specificity, positive or negative predictive value (PPV and NPV) or data from which this could be calculated. Laboratory studies that assessed the validity of new tests or techniques on a sample of galactosaemia patients (with/without comparison to samples from healthy controls) were excluded.

Description of the evidence

Two retrospective screening programme evaluations contained evidence applicable to question 2. One was the 30-year regional evaluation of the Italian NBS study included for question 1[10] which screened 1.1 million newborns in total. This study contained information on total numbers screened, recalled and testing positive allowing calculation of PPV. The second was an evaluation of the effectiveness of the Dutch screening programme[12] which provided test performance data for the 5 methods sequentially used over the first 9 years of the programme, screening in total 1.6 million newborns. The findings from these studies are summarised in Table 5, with full evidence extraction in Appendix 3, Tables 15 and 16. A summary of quality assessment using the QUADAS-2 checklist is given in Table 6, with the detailed appraisal for each study given in Appendix 3, Tables 23 and 24.

To note, the Varela-Lema et al systematic review[4] on screening for classic galactosaemia (2016), discussed in the introductory section of this review, was excluded as evidence for this question. While it reported the test performance findings from 5 NBS evaluations, all were published prior to 2014. As all of these publications were available at the time of the 2015 evidence review, this was not considered to be contributing new evidence.

Table 5. Retrospective NBS programme evaluations providing test performance data

Study, country	Time of screen (mean)	Reference standard	Test	Population	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Porta et al 2015[10] Regional centre Italy, 1982 to 2012	1st screen: 3.4 days Recall for screen +ve: 12.2 days	GALT 0 to 6% plus symptoms (DNA analysis NR)	TGal >10mg/dl (Initial screen and recall) 1982 to 2000: qualitative fluorometric assay 2000 onwards: quantitative fluorescent galactose oxidase method	n=1,123,909 n=8991 1st tier screen +ve (0.8% recall rate) n=33 2nd tier screen +ve: n=13 CG (39%) n=8 DG (24%) n=3 GALK deficiency n=2 other diagnosis n=7 transient raised TGal	NR[1]	NR[1]	0.14% with initial screen positive 39% with repeat screen positive	NR[1]
Welling et al 2017[12] The Netherlands, 2007 to 2015	3 to 7 days	GALT <15% and/or 2 pathogenic variants	Method 1 (2007) TGal ≥700 µmol/l Bio-Rad Quantase Neonatal TGal assay	n=44,174 n=17 positive n=1 TP n=216 FP	100%[2]	99.51%	0.46%	100%[2]
			Method 2 (2007-12) 1st tier: GALT ≤20% 2nd tier: TGal ≥700 µmol/l Bio-Rad CPDA Neonatal GALT assay + Quantase TGal assay	n=952,191 n=322 positive n=18 TP n=304 FP	100%[2]	99.97%	5.6%	100%[2]
			Method 3 (2012-14) 1st tier: GALT ≤15% 2nd tier: TGal ≥700 µmol/l	n=345,685 n=87 positive n=6 TP (+ 1 sibling[3]) n=81 FP	100%[2]	99.98%	6.9%	100%[2]

Study, country	Time of screen (mean)	Reference standard	Test	Population	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
			GALT measure using automated or manual GSP assay					
			Method 4 (2014-15) 1st tier: GALT ≤2.7 units/dl 2nd tier: TGal ≥900 µmol/l (Assay as 3)	n=173,656 n=96 positive n=3 TP (+ 2 siblings[3]) n=93 FP	100%[2]	99.95%	3.1%	100%[2]
			Method 5 (2015) 1st tier: GALT ≤2.0 units/dl 2nd tier: TGal ≥1000 µmol/l (Assay as 3)	n=122,027 n=30 positive n=0 TP n=30 FP	Unknown (no cases)	100%	Unknown	Unknown
			Overall combined	n=1,637,733 n=752 positive n=28 TP (+ 3 siblings[3]) n=724 FP	100%[2]	99.56%	3.6%	100%[2]

Abbreviations: CG, classic galactosaemia; DG, Duarte galactosaemia; FP, false positive; GALT, galactose-1-phosphate uridyltransferase; GSP, Genetic Screening Processor; NR, not reported; PPV, positive predictive value; NPV, negative predictive value; TGal, total galactose; TP, true positive

[1] Porta et al do not confirm there were no screen-negative cases clinically-detected during this period; therefore it is not possible to assume 100% sensitivity and calculate specificity and NPV

[2] Welling et al report there were no cases of CG clinically diagnosed among screen-negatives during the study period; therefore they report 100% sensitivity

[3] 3 cases of CG with affected siblings were treated pre-emptively from birth outside of the screening process; they were not considered to be screen negatives.

Table 6. QUADAS-2 assessment of risk of bias and applicability of diagnostic accuracy studies

Domain	Risk of bias and applicability	Porta et al[10]	Welling et al[12]
Patient selection	Risk of bias	Low	Low
	Applicability concerns	Low	Low
Index test	Risk of bias	Low	Low
	Applicability concerns	Unclear	Unclear
Reference standard	Risk of bias	Unclear	Low
	Applicability concerns	Unclear	Low
Flow and timing	Risk of bias	Unclear	Low

Discussion of findings

These 2 studies are a small total body of evidence, though both are large studies demonstrating classic galactosaemia screening programmes in practise following national implementation.

The Dutch programme evaluation[12] was assessed to be of good quality, with low risk of bias across the 4 QUADAS-2 domains of patient selection, index test conduct, reference standard, flow and timing. The only uncertainty related to applicability of the index test. The tests and thresholds used were continually modified to try and maximise test performance in this population, but it is not possible to know whether these tests and thresholds would perform the same in a potential UK programme. The Italian study[10] had the same low risk of bias for patient selection and index test conduct, with the same uncertainties around applicability of the test and cut-off to other populations. This study had additional uncertainties around the reference standard and flow and timing, though these issues are expected to be due to lack of clarity within the publication, rather than limitations of the screening method itself. These included lack of information over the cut-off threshold used for GALT enzyme activity, and whether or not genetic analysis was used in diagnostic confirmation. Therefore it was not possible to say these diagnoses were compatible with the gold standard. Regarding flow and timing, the publication only gave information for screen-positives. It was not possible to know whether there may have been any diagnoses among those who screened negative at the initial screen, or among screen-positives who then screened negative at their recall test. Therefore test performance statistics other than PPV (sensitivity, specificity and NPV) could not be calculated.

Looking in at the test performance results, both studies screened used different protocols. In the Italian programme[10] infants with raised total galactose on NBS (at 3 days of life) were

recalled, and only upon a repeat confirmation (at 12 days) were they started on galactose-restriction and referred for diagnostic assessment. The Dutch programme[12] continually modified their index test to try and improve test performance, using a total 5 different methods, all taken at 3 to 7 days of life. The initial method in the first year of NBS was a single measurement of raised total galactose (using a different threshold to the Italian programme[10]). This was subsequently changed to a 2-tiered approach, testing firstly GALT enzyme level and then testing total galactose on the same sample if GALT levels were below threshold. Methods 2 to 5 of the Dutch programme[12] altered the thresholds, lowering the GALT cut-off with or without additionally raising the total galactose cut-off, to try and reduce the number of false positives. Screen-positives were referred at median 6 days, and started on galactose-restriction immediately if symptomatic, or at diagnostic confirmation if asymptomatic.

Although their cut-offs and assays were not identical, both the Italian study[10] and method 1 of the Dutch programme[12] demonstrated that a single elevated total galactose in the first days of life had a very low PPV for classic galactosaemia: 0.14% in the Italian programme and 0.46% in the Dutch programme. In the Italian programme[10] if screen-positives were considered to be only those infants who had raised total galactose at the initial screen and at the recall test, then this gives a much higher PPV for classic galactosaemia of 39%. However, this protocol means that 99.6% of infants who were recalled for a repeat test were false positives (excluding those with confirmed galactose elevation) causing potential parental anxiety. The delay of an additional 9 days (screen day 3, recall day 12) would also mean that infants with classic galactosaemia would have considerable galactose exposure, potentially increasing their risk of toxicity (as shown in question 1, three-quarters of cases in this study were readmitted with symptoms before NBS results were available).

The Dutch study[12] demonstrated that measuring GALT activity (≤20%) in addition to total galactose (≥700 µmol/l), increased PPV from 0.46% to 5.6%. In method 3, lowering the GALT cut-off to ≤15% (which is more compatible with the definition of classic galactosaemia) further increased PPV to 6.9%. In method 4, the GALT cut-off was modified to ≤2.7 U/l and the total galactose cut-off increased to ≥900 µmol/l to try and improve specificity. However, this was not demonstrated, as there were a low number of cases in this period and PPV decreased to 3.1%. The programme then further decreased the GALT cut-off and increased the total galactose cut-off in method 5, but it is not possible to assess the effectiveness of this change as there were no cases.

Therefore despite attempts to optimise cut-offs or use different combinations of test, both of these NBS programmes demonstrated screening to have very low PPV for classic galactosaemia. The vast majority of screen-positives were false positives. The Italian study[10] demonstrated that 24% of those with raised total galactose confirmed at the recall

test had Duarte galactosaemia (n=8/33) while 21% had transient raised total galactose (n=7/33), which normalised after a short period on galactose restriction. The Dutch study did not specify any diagnoses among false positives so it is not possible to know the proportion with Duarte or other variants of unknown significance.

However, it is difficult to evaluate the clinical validity of a screening test using PPV alone, because the PPV for rare diseases such as classic galactosaemia will be low even if the test has high sensitivity and specificity. This is further complicated by the index test used and how screen-positives are defined prior to confirmation with the reference standard (for example, in the Italian study, whether this was based on initial or recall screening results). Despite the low PPV, the Dutch programme[12] demonstrated very high specificity, with false positive rate (FPR) varying from 0.49% with method 1 to 0.02% with method 3. There were not known to have been any cases clinically diagnosed since implementation of screening, and therefore sensitivity was reported as 100%. As reported, the Italian centre[10] did not state whether or not any cases were clinically diagnosed among screen-negatives, and so it was not possible to assume this. However, the findings of these studies are overall compatible with that of previous NBS programme evaluations identified by the 2015 evidence summary and the Varela-Lema et al 2016 systematic review.[4] No false negatives have been reported in the published literature to date, indicating that screening programmes for classic galactosaemia have maximum sensitivity. Despite some variability, all studies to-date have also demonstrated PPV to be overall low, with a substantial proportion of false positives reported to have Duarte galactosaemia. Considering also the variable index tests used in previous screening programme evaluations, there does not appear to be a clearly established optimal screening test and cut-off to use for classic galactosaemia screening.

Summary of Findings Relevant to Criteria 4 and 5: Criteria not met†

Two national newborn screening programme evaluations were identified, from Italy and The Netherlands. These 2 large studies were overall of good quality, with the main uncertainty being the applicability of the index tests and cut-offs used to the UK.

† **Met** -for example, this should be applied in circumstances in which there is a sufficient volume of evidence of sufficient quality to judge an outcome or effect which is unlikely to be changed by further research or systematic review.
Not Met - for example, this should be applied in circumstances where there is insufficient evidence to clearly judge an outcome or effect or where there is sufficient evidence of poor performance.
Uncertain -for example, this should be applied in circumstances in which the constraints of an evidence summary prevent a reliable answer to the question. An example of this may be when the need for a systematic review and meta-analysis is identified by the rapid review.

In the Italian programme, screening involved an initial and repeat total galactose measurement; in the Dutch programme the index tests (total galactose with/without GALT enzyme activity) and cut-offs were continually modified to try and optimise test performance. Both studies demonstrated that all tests had very poor PPV for classic galactosaemia, ranging from 0.14% in the Italian programme to a maximum of 6.9% in the Dutch programme. Therefore the majority of screen positives in these screening programmes were false positives. The Italian programme demonstrated that a quarter of screen positives had Duarte galactosaemia, which does not require management or monitoring (the Dutch programme did detail any diagnoses among false positives). Despite the low PPV (which will be influenced by low population prevalence) the Dutch programme demonstrated that the tests had 100% sensitivity for identifying newborns with classic galactosaemia with no clinically-diagnosed cases among false negatives. Specificity was also very high with FPR ranging from 0.49% to 0.02% depending on the test and method used (the Italian programme did not have data available to calculate sensitivity or specificity). These findings of maximum sensitivity and high specificity but generally poor PPV are overall compatible with previous screening programme evaluations. Neither does there appear to be a clearly established optimal screening test and cut-off to use for classic galactosaemia screening.

Though further evidence on the sensitivity and specificity of tests to detect classic galactosaemia would be ideal, it is acknowledged that this is difficult in the study of rare diseases. Finding ways to address that is important, particularly given the potential for identification of Duarte galactosaemia and other variants of unknown clinical significance. Further screening studies with improved methodological consistency (in terms of index test cut-offs, repeat testing and reference standard) may be achievable and would allow for an informative evaluation of a test to be used in newborn screening for classic galactosaemia.

Criterion 9 — Early initiation of treatment

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"

Question 3 – Does early initiation of treatment for individuals with classic galactosaemia provide better short- and long-term outcomes?

The 2015 UK NSC evidence summary[1] found limited and low quality evidence comparing short- and long-term outcomes between screen-detected and clinically-detected cases. A

single before-after study from Germany reported that screening reduced mortality and improved longer-term outcomes compared with clinical detection, but there was no statistical comparison and no further delineation of results. Other retrospective cohort studies/case series conversely reported that screen-detection and subsequent treatment was not associated with long-term outcomes, but no further details were available.

Prior to this, the 1997 HTA[3] had found conflicting evidence. One US study had reported that mortality due to sepsis was higher when diagnosis and treatment were delayed to the second week of life. However, the only UK study found that screening had no effect on acute neonatal illness compared with no screening.

This evidence review aimed to evaluate whether early initiation of treatment for individuals with classic galactosaemia provides better short- and long-term outcomes.

Eligibility for inclusion in the review

The reviewers aimed to identify studies that reported on the effectiveness of early initiation of galactose-restriction (following screen-detection or otherwise) compared with later treatment initiation (non-screening context or other comparison of later treatment) on short- and long-term outcomes. The ideal study designs would be randomised controlled trials (for example, screening compared to no screening) followed by prospective comparative cohort studies. If these study designs were not available, the reviewers aimed to look at any retrospective cohort studies or case series that provided any comparison of outcomes between individuals receiving earlier treatment and those receiving a later treatment.

The short-term outcomes were considered to be neonatal complications and mortality. Long-term outcomes could include neurodevelopmental outcomes, bone health, cataracts, reproductive complications and quality of life. Studies from the UK or other representative western/European countries were eligible.

Description of the evidence

Of the 22 publications retrieved at full text across the general topic search, 3 studies provided evidence for this question: 1 systematic review of randomised controlled trials (RCTs)[13] and 2 case series.[14, 15] No randomised controlled trials or prospective comparative cohort studies were identified.

The Cochrane systematic review (search date end 2017)[13] had searched for RCTs comparing screening vs no screening and looking at outcomes of mortality or disease morbidity. No trials met inclusion criteria. Of the 2 case series, one was the evaluation of

patients in the GalNet registry[15] (described in the introductory section), which as of end 2018 had covered 15 countries (32 centres) and 509 patients. This study provided some comparison of short- and long-term outcomes between screened and non-screened individuals, or earlier compared with later initiation of treatment. The second case series[14] reviewed childhood neurodevelopmental outcomes for patients homozygous for the common pathogenic variant c.563A>G (p.Gln188Arg) who were seen over a 14-year period in a single Turkish centre. As reported, 58% of cases recorded in the GalNet registry[15] to date have this genotype. The study included comparison of outcomes according to age at diagnosis.

The findings from these studies are summarised in Table 7, with full evidence extraction in Appendix 3, Table 17. Quality assessment is summarised in Table 8, with the detailed appraisal for each study given in Appendix 3, Tables 25 to 27. The reviewers assessed the 2 case series using CASP checklist for cohort studies. Notably, these assessments were based on the quality and applicability of the study content to address the comparative key question; they do not necessarily represent the quality of these studies for its intended purpose as a descriptive case series.

The retrospective case series and screening programme evaluations included for questions 1 and 2 were also reviewed to see if they contained any information applicable to this question. The NBS evaluations by Porta et al[10] (included for questions 1 and 2) and Welling et al[12] (question 2) provided no comparison between individuals screened vs non-screened or treated early vs late. However, they provided input to the discussion on clinical outcomes following screening. They are additionally discussed below but do not provide direct evidence applicable to this question. The 4 remaining case series included for question 1 were not considered to provide information relevant to this question (analysis of their potential applicability is detailed in the data extraction tables for these studies in Appendix 3, Table 15).

The Varela-Lema et al systematic review of screening for classic galactosaemia[4] discussed in the introductory section, was excluded as evidence for this question. While this systematic review covered information on outcomes in relation to screening or timing of treatment, all of the analysed studies were published prior to 2014, so were available either at the time of the 2015 evidence review or earlier 1997 HTA,[3] and were not considered to be contributing new evidence.

Table 7. Studies providing data on the association between timing of treatment initiation and outcomes

Study, design	Population	Intervention	Comparator	Outcome
Lak et al 2017[13] Systematic review, search to Dec 2017	Eligible: RCTs of NBS screening in 1st week of life via bloodspot or venous sample vs no screening. No studies identified.	NA	NA	NA
Rubio-Gozalbo et al 2019[15] Case series International registry, 15 countries, data collection 2014 (inception) to date	Total n=509 with GALT activity ≤10% and/or pathogenic variants Median age 18 years 48% male, 94% white No breakdown by comparison groups	NBS n=215/468 (46%) (no further detail)	No NBS n=253/468 (54%)	Risk of any neonatal complications: OR 0.30, 95% CI 0.20 to 0.47 (p<0.0000001) Risk of any neurological complications: OR 0.32, 95% CI 0.20 to 0.51 (p<0.00001) No absolute numbers or outcomes listed for those screened vs not Patients diagnosed following NBS were reportedly 'often younger (p<0.000001) and started diet in the 1st week of life (p<0.000000000001)'
		Galactose-restriction started in 1st week life n=198/391 (51%)	Galactose-restriction started after 1st week n=193/391 (49%)	Risk of neonatal complications: OR 0.32, 95% CI 0.21 to 0.50 (p<0.000001) No absolute numbers or outcomes listed for those treated early vs late
Ozgun et al 2019[14] Case series Single centre, Turkey 2003 to 2017	n=46 homozygous for p.Gln188Arg and with complete assessments available Median age at latest assessments: 3 to 4 years	Age at diagnosis: n=20/46 (43%) diagnosed ≤1 week (1 to 7 days)	Age at diagnosis: n=26/46 diagnosed >1 week: n=10 in 8 to 14 days n=7 in 15 to 21 days n=4 in 22 to 28 days n=2 in 29 to 35 days n=3 in 36 to 42 days	Developmental delay: Age ≥20% lower than chronological age on Denver II screen: n=21 (46%) vs n=25 (54%) normal development Age at diagnosis: 2.38 ± 1.75 weeks vs 2.20 ± 1.35 weeks normal (p=0.954) Neurological examination: n=11 (24%) with abnormalities (ataxia in 55%, n=6) vs 35 (76%) with normal examination Age at diagnosis: 2.9 ± 1.7 weeks vs 2.0 ± 1.4 weeks normal (p=0.137) Frequency of problems by age at diagnosis:

				• Number with developmental delay: n=10/20 (50%) ≤1 week vs n=11/26 (42%) >1 week (p=0.796) • Abnormal neurological exam: n=3/20 (15%) vs n=8/26 (31%) (p=0.187) • Pathology on brain MRI: n=10/20 (50%) vs n=12/26 (46%) (p=0.515) • Convulsion (no detail): n=6/20 (30%) vs n=5/26 (19%) (p=0.307) Spearman's test: no correlation between time of diagnosis and developmental delay (p=0.954), abnormal neurological examination (p=0.137), or presence of MRI pathology (p=0.917)

Abbreviations: CI, **confidence interval;** GALT, galactose-1-phosphate uridyltransferase; NA, not applicable; **NBS newborn screening; OR odds ratio**; RCT, randomised controlled trial

Table 8. Quality assessment using the CASP cohort study checklist

Domain	Assessment	Risk of bias	
		Rubio-Gozalbo et al[15]	Ozgun et al[14]
Study validity	Focused question addressed	High[1]	High[1]
	Non-selective recruitment	Unclear	Unclear
	Reliable exposure assessment	Unclear	Unclear
	Reliable outcome assessment	Unclear	Low
	Potential confounders assessed	High	Low
	Complete follow-up	Unclear	Unclear
Results	Clear results (absolute and relative)	High	Unclear
	Precision of results	Unclear	Unclear
Applicability	Applicable to local population	Unclear	Unclear
	Possible to draw implications for practice	Unclear	Unclear

[1] Assessed as high risk because these were not designed as comparative studies

Discussion of findings

As demonstrated by the Cochrane systematic review[13] no RCTs have been conducted that have compared short- or long-term outcomes for screened compared with non-screened individuals. Neither did this evidence update identify any prospective cohorts that have compared outcomes for children by screening status or time of treatment initiation.

The 2 identified case series provide a small body of inconsistent evidence. In the international GalNet registry[15] roughly half of all registered were respectively identified through NBS, or started galactose-restriction in the first week of life. This study found that NBS was associated with around a 70% reduction in relative risk of neonatal complications or long-term neurological complications (unclear age at assessment, sample median 18 years). Starting galactose-restriction in the first week of life was similarly associated with around a 70% risk reduction. The Turkish case series[14] conversely found no association between age at diagnosis and the risk of developmental delay, abnormality on neurological examination, or pathology on brain MRI (assessments at around 3 to 4 years of age).

The 2 case series were overall assessed to be low quality as comparative studies due to gaps in the data and several uncertainties or lack of clarity in reporting. Both also had applicability issues. However, this was for different reasons, necessitating separate discussion for each study.

The GalNet registry study[15] has strengths in its international representation and large sample size (n=509) which should make the analyses more reliable. However, this was not designed as a comparative study to examine the effect of the 2 exposures of NBS or time of treatment on outcomes and there are several uncertainties and gaps in the data. All disease characteristics, including the prevalence of short- or long-term complications are given for the cohort as a whole. Absolute rates within the 2 comparison groups are not given. The registry was based on submitted records and was not fully comprehensive for all exposure and outcomes. The majority of the cohort (468/509) had information available on NBS status, but time of treatment initiation was only documented for 77% (394/509). For outcomes, the presence or absence of neonatal complications was documented for 82% (416/509) while neurological assessments were documented for only 63% (320/509). Without the absolute numbers, it is difficult to know how many individuals within the respective comparison groups (screened vs non-screened, early vs late treatment) had this outcome data available. Therefore the data on which these statistical comparisons were based is very unclear. It is also possible that patients who had assessment data available were not representative, as those with potential complications (such as neurological) may have been more likely to have had assessments than others.

There was minimal comparative assessment for long-term outcomes. The study analysed the link between NBS and neurological complications, but did not look at the association with others such as developmental delay, language and speech disorders, mental and behavioural problems. Time of dietary implementation was assessed only in association with neonatal complications, with no analysis for long-term complications. There is, however, uncertainty as to how much overlap there may have been between the 46% of the cohort who received NBS and the 51% who started treatment within the first week of life. The study reports that those screened were more likely to start treatment early, but it is unclear whether they were essentially the same group (possibly accounting for the similar risk reduction for neonatal complications).

Confounding is potential source of bias. The GalNet registry includes a heterogeneous mix of cases and it may have been relevant to adjust the analysis for the potential influence of variables such as current age/birth cohort, country of origin, genotype and enzyme activity (though notably the vast majority had absent/profoundly deficient activity). Lastly, although this is a large international sample, the estimated coverage of the GalNet registry in terms of all European/Western people with classic galactosaemia is unclear. For example, up to 2018 the registry covered 13 European (including UK), US and Israel, and 32 centres within these countries. Some countries are not covered (such as Italy, Canada, Scandinavia and Australasia), while for included countries, it is unclear whether all centres caring for those with the condition are covered (for example, only one US centre is included). This extensive range of uncertainties result in this being low quality evidence to inform whether screening or early treatment may improve outcomes in classic galactosaemia.

The Turkish case series[14] by comparison was a smaller study but benefited from being a homogenous sample, with all cases of similar age, the same genotype, nationality and locality. Analysing those of the same genotype is particularly relevant, as in a variable sample of homozygotes or compound heterozygotes, the severity of the variants may influence both the time of presentation/diagnosis/treatment and the likelihood of complications. This made this study at low risk from confounding bias and better able look at the direct effect of time of treatment upon outcomes. All cases also had complete outcome data available. However, the fact that selection was on this basis may have adversely affected representation. The authors report that 68 patients were seen at this centre over the study period, 68% of whom (46) were eligible for this analysis. It is unclear whether the 68 included all cases with classic galactosaemia in this centre, or all homozygotes for p.Gln188Arg. As such it is difficult to know how many with this genotype had complete outcome data. Also, similar to the GalNet registry, it is possible that patients experiencing neurodevelopmental problems may have been more likely to have had assessments than those without. There were also uncertainties with the exposure which was defined only as ‘time of diagnosis.’ It is unclear whether this was compatible with the

time of treatment initiation, and whether it referred to initial clinical suspicion/presentation or diagnostic confirmation. The small sample size also limits the statistical analysis, with the possibility that a true effect of early treatment may have been missed. Lastly the findings from this homozygous Turkish sample may not be applicable to patients with the same genotype in the UK, where care approaches may differ.

These 2 studies were the only sources of comparative data. Of the 2 NBS studies analysed for question 2, the Italian regional evaluation reported that the majority of cases identified through NBS (10/13, 77%) had mild to severe intellectual disability. One interpretation of this could be that NBS did not prevent complications. However, the study did report that 3 siblings of those with severe variants who were put on galactose restriction 'immediately after DBS collection' had earlier normalisation of galactose and an uncomplicated clinical course (mild jaundice only), in contrast with the 'life-threatening clinical course' of the older sibling. This is similar to previous research which has suggested that pre-emptive treatment of siblings from birth might prevent complications. Pre-emptive treatment would clearly not be possible on a universal basis, though the wording of this study appears to suggest that treatment does not have to be from birth, but starting from the time of the dried blood spot (performed in the study at mean 3.4 days of life) can still improve outcomes. However, this is a tenuous interpretation from this description; it is unclear why treatment should be delayed to the time of the DBS in these cases rather than be immediate from birth to await assessment of enzyme activity/genotype. The Netherlands evaluation did not perform long-term follow-up but reported that all screening methods 'prevented critical illness and death.' However, without comparison it is not possible to draw any further interpretation over the effect of NBS and time of treatment initiation upon outcomes.

Summary of Findings Relevant to Criterion 9: Criteria not met‡

This evidence summary update found a limited body of low quality evidence to inform whether early initiation of treatment improves short- and long-term outcomes in classic galactosaemia. No randomised controlled trials have been conducted that have compared outcomes for screened compared with non-screened individuals (where screening may be a marker for earlier initiation of treatment). Likewise no prospective comparative cohorts have compared outcomes for children by screening status or time of treatment initiation.

Only 2 cases series provided evidence for this question. These were the international GalNet registry, which since 2014 has recorded data on 509 patients with classic galactosaemia from 15 countries, and a Turkish case series of 46 patients homozygous for the common pathogenic variant p.Gln188Arg. Both were assessed to be low-quality as comparative studies, with various uncertainties and gaps in the data. For example, both studies had the potential for selection bias (such as only including those who had complete assessments documented), the analysis in the GalNet registry lacked absolute numbers and had the potential for confounding (for example, by genotype, country of origin, care provision), and the Turkish study was limited by small sample size and uncertain applicability to the UK. The 2 studies had inconsistent findings. The GalNet registry found that newborn screening and galactose-restriction in the first week of life were both associated with a 70% reduced relative risk of short-term neonatal complications. Newborn screening was also associated with a 70% reduced risk of longer-term neurological outcomes (long-term outcomes were not assessed for early treatment). The Turkish case series[14] conversely found no association between age at diagnosis and risk of longer-term childhood developmental delay or neurological problems.

Overall the findings of this evidence update reflect past studies which have found a limited quantity of low-quality evidence which has given mixed findings on whether or not early treatment improves outcomes in classic galactosaemia.

‡ **Met** -for example, this should be applied in circumstances in which there is a sufficient volume of evidence of sufficient quality to judge an outcome or effect which is unlikely to be changed by further research or systematic review.
Not Met - for example, this should be applied in circumstances where there is insufficient evidence to clearly judge an outcome or effect or where there is sufficient evidence of poor performance.
Uncertain -for example, this should be applied in circumstances in which the constraints of an evidence summary prevent a reliable answer to the question. An example of this may be when the need for a systematic review and meta-analysis is identified by the rapid review.

Review summary

Conclusions and implications for policy

Based on the overall synthesis of evidence published since the last UK NSC review in 2015, newborn screening for classic galactosaemia is still not recommended.

Three key questions were considered in this rapid review to address whether there has been a development in the evidence base relating to (1) the age of presentation of classic galactosaemia, (2) an accurate screening test for the identification of newborns with classic galactosaemia, and (3) demonstration of a benefit of early initiation of treatment following screening. Nine studies were deemed to be applicable to the key questions.

Median age at presentation
This evidence summary update identified 5 case series providing information on the average age of presentation in classic galactosaemia, 3 from countries where newborn screening is not performed (UK,[7] Croatia[8] and Turkey[9]), and 2 Italian studies[10, 11] evaluating cases identified through from newborn screening programmes.

The studies represented different settings, and the quality varied, with some lacking clarity on the diagnostic criteria used or the comprehensive of representation. Nevertheless the findings were overall consistent and suggest that most infants will present clinically by around 7 days of life. Signs of liver dysfunction/ liver failure appear to be universal as the presenting symptom, with sepsis, cataracts and lethargy also frequently reported. The 2 Italian newborn screening studies reported that the majority of cases (>75%) presented clinically before screening results were received or a diagnosis made.

The findings suggest the timing of development from latent to declared disease in classic galactosaemia is reasonably understood. However, this may raise questions over the benefits of screening given that it may not be able to prevent neonatal toxicity. It is expected that a large proportion of screen-detected cases would be symptomatic by the time screen results are confirmed and diagnosis is made.

Screening test performance
Two retrospective, national newborn screening programme evaluations were identified, from Italy[10] and The Netherlands.[12] These 2 large studies were overall of good quality, with the main uncertainty being the applicability of the index tests and cut-offs to the UK or other populations.

In the Italian programme[10] screening involved an initial and repeat total galactose measurement; in the Dutch programme[12] the index tests (total galactose with/without GALT enzyme activity) and cut-offs were continually modified to try and optimise test performance. Both studies demonstrated that all tests had very poor positive predictive value (PPV) for classic galactosaemia, ranging from 0.14% in the Italian programme[10] to a maximum of 6.9% in the Dutch programme.[12] Therefore the majority of screen positives in these screening programmes were false positives. The Italian programme[10] demonstrated that a quarter of screen positives had Duarte galactosaemia, which does not require management or monitoring (the Dutch programme[12] did detail any diagnoses among false positives). Despite the low PPV (which will be influenced by low population prevalence) the Dutch programme[12] demonstrated that the tests had 100% sensitivity for identifying newborns with classic galactosaemia, with no clinically-diagnosed cases among false negatives. Specificity was also very high with a low false positive rate ranging from 0.49% to 0.02% depending on the test and method used (the Italian programme[10] did not have data available on screen negatives to calculate sensitivity or specificity).

These findings of maximum sensitivity and high specificity but generally poor PPV, with the potential for identification of Duarte galactosaemia and other variants of unknown clinical significance, are overall compatible with previous screening programme evaluations. On this basis, there was insufficient evidence to establish an optimal screening test and cut-off to use for classic galactosaemia screening.

Early initiation of galactose-restriction

No randomised controlled trials have been conducted that have compared outcomes for screened compared with non-screened individuals (where screening may be a marker for earlier initiation of treatment), as confirmed by one 2017 Cochrane systematic review.[13] Likewise no prospective comparative cohorts have been conducted that have compared outcomes for children by screening status or time of treatment initiation.

Only 2 case series[14, 15] provided evidence for this question. These were the international GalNet registry,[15] which since 2014 has recorded data on 509 patients with classic galactosaemia from 15 countries, and a Turkish case series[14] of 46 patients homozygous for the common pathogenic variant p.Gln188Arg. Both studies were assessed to be low-quality as comparative studies, with various uncertainties and gaps in the data. For example, both studies had the potential for selection bias (such as only including those who had complete assessments documented), the analysis in the GalNet registry lacked absolute numbers and had the potential for confounding (for example, by genotype, country of origin, care provision), and the Turkish study was limited by small sample size and uncertain applicability to the UK. The 2 studies had inconsistent findings. The GalNet registry found that newborn screening and galactose-restriction in the first week of life were

both associated with a 70% reduced relative risk of short-term neonatal complications. Newborn screening was also associated with a 70% reduced risk of longer-term neurological outcomes (long-term outcomes were not assessed for early treatment). The Turkish case series[14] conversely found no association between age at diagnosis and risk of longer-term childhood developmental delay or neurological problems.

Overall, the findings of this UK NSC evidence summary are in agreement with the 2015 UK NSC evidence summary, finding a limited quantity of low-quality evidence, which has given mixed findings on the association between early treatment and long-term outcomes in classic galactosaemia.

Limitations

This UK NSC evidence summary was conducted using a rapid review methodology. The searches were limited to 3 literature databases and did not include grey literature resources. This evidence review included only peer-reviewed journal publications in the English language. The reviewers were also unable to contact study authors or review non-published material. However, this is an accepted methodological adjustment for a rapid review, and these limitations should not have led to the exclusion of any pivotal studies.

Appendix 1 — Search strategy

Electronic databases

The search strategy included searches of the databases shown in Table 9. The initial searches were conducted in the Cochrane Library and Embase. Due to the small body of evidence, the search was additionally run in Pubmed to see whether this contained any relevant citations not covered by Embase.

Table 9. Summary of electronic database searches and dates

Database	Platform	Searched on date	Date range of search
Embase	Embase.com	22 January 2020	2014 - 2020
The Cochrane Library, including: **Cochrane Database of Systematic Reviews (CDSR)** **Cochrane Central Register of Controlled Trials (CENTRAL)**	Wiley Online	22 January 2020	2014 - 2020
PubMed	pubmed.gov	18 February 2020	2014 - 2020

Search Terms

Search terms included combinations of free text and subject headings (Medical Subject Headings [MeSH] for MEDLINE, and Emtree terms for Embase), around the disease area: galactosaemia.

Search terms for Embase are shown in Table 10, search terms for the Cochrane Library databases are shown in Table 11, and search terms for PubMed are shown in Table 12.

Table 10. Search strategy for Embase

Term Group	#	Search terms	Results
Condition	#1	galactosemia'/exp	3240
Condition	#2	galactosaemia OR galactosemia	3721
Condition	#3	galactose 1 phosphate uridylyltransferase'/exp	948
Condition	#4	galt NEAR/5 deficien*	188

Condition	#5	galactose 1 phosphate uridylyltransferase' OR 'galactose 1 phosphate uridylyl transferase' OR 'galactose-1-phosphate uridyl-transferase' OR 'galactose-1-phosphate uridyltransferase' 'galactose phosphate uridyl transferase' OR 'galactose phosphate uridyltransferase'	1189
Condition	#6	1 OR #2 OR #3 OR #4 OR #5 AND [english]/lim AND [2014-2020]/py	679
Animal studies	#7	1 OR #2 OR #3 OR #4 OR #5 AND [english]/lim AND [2014-2020]/py AND ('animal cell'/de OR 'animal experiment'/de OR 'animal model'/de OR 'animal tissue'/de OR 'in vitro study'/de OR 'nonhuman'/de)	129
Remove animal studies	#8	#6 NOT #7	550
Conference abstracts	#9	#6 NOT #7 AND [conference abstract]/lim	193
Remove conference abstracts	#10	#8 NOT #9	361

Table 11. Search strategy for The Cochrane Library

Term Group	#	Search terms	Results
Condition	#1	MeSH descriptor: [Galactosemias] explode all trees	9
Condition	#2	galactosaemia* OR galactosemia*	25
Condition	#3	galt NEAR/5 deficien*	1
Condition	#4	"galactose 1 phosphate uridylyltransferase" OR "galactose 1 phosphate uridylyl transferase" OR "galactose-1-phosphate uridyl-transferase" OR "galactose-1-phosphate uridyltransferase" OR "galactose phosphate uridyl transferase" OR "galactose phosphate uridyltransferase"	3
Condition	#5	(or #1-#4)	25
Limit to Reviews only with date restriction FINAL 1	#6	#5 with Cochrane Library publication date Between Jan 2014 and Jan 2020, in Cochrane Reviews	3
Limit to Central only with date restriction	#7	#5 with Publication Year from 2014 to 2020, in Trials	7

Table 12. Search strategy for PubMed

Term Group	#	Search terms	Results
Condition	#1	"Galactosemias"[Mesh]	2248

Condition	#2	galactosemi* or galactosaemi*	2836
Condition	#3	GALT deficien*	102
Condition	#4	(Galactose-1-Phosphate Uridyl-Transferase Deficien*) OR (Galactose 1 Phosphate Uridyl Transferase Deficien*) OR (Galactose-1-Phosphate Uridyltransferase Deficien*) OR (Galactose 1 Phosphate Uridyltransferase Deficien*)	55
Condition	#5	#1 or #2 or #3 or #4	2846
Limit to English language, date restriction, and Humans only	#6	#5 Publication date from 2014/01/01 to 2020/12/31; Humans; English	165

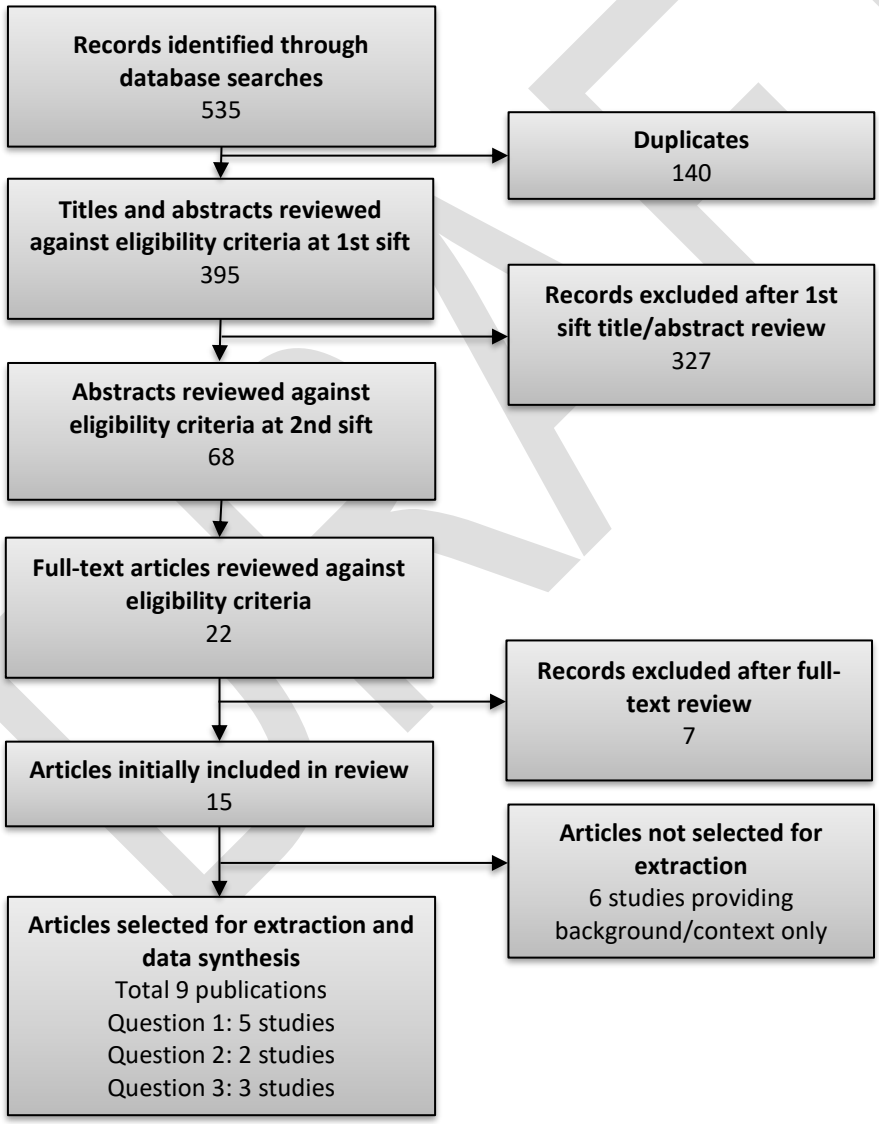
Results were imported into EndNote and de-duplicated.

Appendix 2 — Included and excluded studies

PRISMA flowchart

Figure 2 summarises the volume of publications included and excluded at each stage of the review. Nine publications were ultimately judged to be relevant to one or more review questions and were considered for extraction. Publications that were included or excluded after the review of full-text articles are detailed below.

Figure 2. Summary of publications included and excluded at each stage of the review



Publications included after review of full-text articles

The 9 publications included after review of full-texts are summarised in Table 13 below. Studies were prioritised for extraction and data synthesis. It was planned *a priori* that the tiered approach would be taken to prioritise studies for extraction:

1. Systematic reviews and meta-analyses of applicable study designs would be considered the highest quality of evidence if any were identified. Following this, study designs would be prioritised for each question in the order listed in Table 2.
2. Studies would be prioritised if they considered a UK population, followed by studies from Western populations comparable to the UK.

Publications not selected for extraction and data synthesis are detailed in Table 14.

Table 13. Summary of publications included for each key question

Study	The condition	The test	The intervention	Implementation criteria
Hegarty et al 2015[7]	Q1	-	-	1
Porta et al 2015[10]	Q1	Q2		1, 4,5
Ramadza et al 2018[8]	Q1	-	-	1
Teke Kisa et al 2019[9]	Q1	-	-	1
Viggiano et al 2015[11]	Q1	-	-	1
Welling et al 2017[12]	-	Q2	-	4,5
Lak et al 2017[13]	-	-	Q3	9
Ozgun et al[14]	-	-	Q3	9
Rubio-Gozalbo et al 2019[15]	-	-	Q3	9

Publications excluded after review of full-text articles

Of the 22 publications included after the review of titles and abstracts, 6 articles provided background information only but did not contain information directly relevant to any of the key questions. A further 7 articles were ultimately judged not to be relevant to this review. These publications, along with reasons for exclusion, are listed in Table 14.

Table 14. Publications excluded for question applicability at full-text review

Reference	Reason for exclusion for key question
Excluded publications (providing background information only)	
Pasquali M, Yu C, Coffee B. Laboratory diagnosis of galactosemia: a technical standard and guideline of the American College of Medical Genetics and Genomics (ACMG). Genet Med. 2018;20(1):3-11.	American College of Medical Genetics and Genomics (ACMG) Technical Standards and Guidelines for Clinical Genetics Laboratories. Useful background to galactose metabolism pathway and the different types of galactosaemia. Covers requirements for laboratory analysis, storage of samples etc. No question applicability.
Van Calcar SC, Bernstein LE, Rohr FJ, et al. A re-evaluation of life-long severe galactose restriction for the nutrition management of classic galactosemia. Molecular Genetics and Metabolism. 2014;112(3):191-7.	US task force group supplemented by literature review. Reviewing galactose content of different foods and forming consensus/evidence based recommendations on dietary recommendations. No coverage of outcomes of early vs no/late dietary restriction for question 3. Useful background but question applicability.
Varela-Lema L, Paz-Valinas L, Atienza-Merino G, et al. Appropriateness of newborn screening for classic galactosaemia: a systematic review. Journal of Inherited Metabolic Disease. 2016;39(5):633-49.	Spanish review essentially assessing whether screening fulfils criteria (ie. essentially similar task as this evidence update). Involves 2 SRs (search 2014): one for disease information and the second on screening test accuracy. Disease aetiology/management: Said to be no population-based cohorts with prospective follow-up; identifies n=45 studies assessing the disease were small cases series/cross sectional analyses. Useful background information on the aetiology and management, but aside from Viggiano 2015 and van Calcar 2014 (separately reviewed in this update) all other discussed studies pre-date 2013 therefore are not contributing new evidence. Accuracy: n=22 studies reported on 'existing screening methods', n=14 with information on NBS programmes from regional/national reports. Majority said to give descriptive analysis of cases but not to provide screening accuracy results, aside from 5 giving test performance data. These 5 studies date from <2013 and descriptive results only (no meta-analysis).

	Overall without meta-analysis, all test performance studies were available at the time of the last NSC review/HTA and are not considered to be providing new evidence. Background to the condition and to testing, but no content applicable to question.
Viggiano E, Marabotti A, Politano L, et al. Galactose-1-phosphate uridyltransferase deficiency: A literature review of the putative mechanisms of short and long-term complications and allelic variants. Clin Genet. 2018;93(2):206-15.	Narrative on the potential biochemical mechanisms behind symptoms and association between genotype and phenotype. Useful background but no question applicability
Villoria JG, Pajares S, López RM, et al. Neonatal Screening for Inherited Metabolic Diseases in 2016. Seminars in Pediatric Neurology. 2016;23(4):257-72.	Reviewed at full text to see if the study provided information on screening worldwide and anything on cases identified through PKU screening. Contains a table on the number of countries/states that screen for classic galactosaemia but minimal additional information. Any discussed background related to the key questions (e.g. early vs late treatment) is referenced to pre-2013 publications.
Welling L, Bernstein LE, Berry GT, et al. International clinical guideline for the management of classical galactosemia: diagnosis, treatment, and follow-up. Journal of Inherited Metabolic Disease. 2017;40(2):171-6.	Systematic review by guideline working group forming graded recommendations. Applicable to CG as defined by enzyme activity <15%. Covers recommendations: Diagnostic gold standard (enzyme activity in RBCs and/or by gene analysis only) Outlines Duarte – treatment not recommended Treating those with enzyme activity <10% and with pathological variants on both alleles of *GALT* gene Dietary recommendations Follow-up assessments through childhood Background but no question applicability
Excluded publications (not cited within the review)	
Cohen AS, Baurek M et al. Including Classical Galactosaemia in the Expanded Newborn Screening Panel Using Tandem Mass Spectrometry for Galactose-1-Phosphate. International Journal of Neonatal Screening. 2019; 5 (9).doi:10.3390/ijns5020019	Exclude. Case control to validate a new method to measure hexose mono-phosphates (HMP) by tandem mass spectrometry on premise that elevated GLA-1-P will cause HMP to rise. This method aims to overcome problem of not being able to differentiate between HMPs by selectively depleting interfering HMPs using a mild hydrazine solution to improve sensitivity. Included n=5500 healthy samples, n=14 confirmed positive for galactosaemia (2009-15) and n=10 with other metabolic samples. Only gives histograms on distributions of GAL-1-P concentrations and GALT enzyme activity across samples. No test performance data. Not applicable to test performance data.
Frederick AB, Cutler DJ, Fridovich-Keil JL. Rigor of non-dairy galactose restriction in early childhood,	Exclude. Retrospective study of diet and outcomes for n=231 cases selected from the ongoing US cohort 'Bases of pathophysiology and modifiers of

measured by retrospective survey, does not associate with severity of five long-term outcomes quantified in 231 children and adults with classic galactosemia. Journal of Inherited Metabolic Disease. 2017;40(6):813-2	outcome in galactosaemia' running since 1992. Dietary information collected by parental survey from birth to 6 years. Looked at association between the level of dietary restriction and selected childhood developmental/educational outcomes (finding no clear link). Excluded as no content applicable to question of early vs no/late intervention.
Hermans ME, Welsink-Karssies MM, Bosch AM, et al. Cognitive functioning in patients with classical galactosemia: A systematic review. Orphanet Journal of Rare Diseases. 2019;14(1).	Exclude. SR (search Oct 18) to assess literature on the incidence of cognitive impairment and which cognitive domains are impaired. N=11 studies, n=177 cases including child and adult. Narrative discussion of the findings with limited quantitative information aside from tabulation of z scores for results of different functioning tests performed in individual studies. No content applicable to questions on presentation age or association between outcomes and timing of treatment.
Morell-Garcia D, Bauça JM, Barceló A, et al. Usefulness of Benedict's test for the screening of galactosemia. Clinical Biochemistry. 2014;47(9):857-9.	Exclude. Spain. Retrospective analysis of n=159 patients with suspected IEMs screened using Benedict's test (rapid analysis of reducing sugars in urine). Reports 27 tests (17%) positive, a quarter of which had to be discarded for glycosuria. No cases of galactosaemia, no test performance data. (All cited studies in discussion related to galactosaemia screen test performance date pre-2013).
Stroek K, Bouva MJ, Schielen PCJI, et al. Recommendations for newborn screening for galactokinase deficiency: A systematic review and evaluation of Dutch newborn screening data. Molecular Genetics and Metabolism. 2018;124(1):50-6.	Exclude. No relevance to topic. SR on screening for galactokinase (GALK) deficiency. Checked at full text as abstract reported 4 studies screening for GALK deficiency in combination with GALT activity measurement. Porta (2015) has been included; 3 other studies (US and Germany) date pre-2013 though in any case do not contain relevant test performance data on GALT deficiency.
Therrell BL, Lloyd-Puryear MA, Camp KM, et al. 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. 2014;113(1):14-26.	Exclude. No content specific to galactosaemia (uncertain from abstract). Gives overall incidence of IEMs, costs associated with nutritional intervention products, and corporate policies.
Yuzyuk T, Viau K, Andrews A, et al. Biochemical changes and clinical outcomes in 34 patients with classic galactosemia. Journal of Inherited Metabolic Disease. 2018;41(2):197-208.	Exclude. N=16 cases diagnosed through NBS and treated at the Metabolic Clinic of the University of Utah. Children were divided by GAL-1-P levels and clinical outcomes were compared the between groups; but no comparison of early vs no/late treatment. No question applicability.

Appendix 3 — Summary and appraisal of individual studies

Data Extraction

Table 15. Studies relevant to key question 1

Study reference Question	Study design	Population	Method of screen detection or diagnosis	Clinical presentation: age and symptoms	Outcomes and any relation to time of diagnosis/management
Hegarty et al 2015[7]	Retrospective case series of children presenting with acute liver failure (ALF) at King's College London Hospital, UK January 2001 to December 2011 (Non-screening context)	n=127 children <5 years of age with ALF defined as: • no known evidence of chronic liver disease • biochemical evidence of acute liver injury; and • hepatic-based coagulopathy (prothrombin time ≥15 s or international normalised ratio ≥1.5 not corrected by Vitamin K in the presence of hepatic encephalopathy; or a PT ≥20 s or INR ≥2.0 alone) n=36 (17 male) with underlying inherited metabolic disease (IMD) following metabolic investigations including	Measurement of GALT level (cut-off not given) No further detail or genetic analysis given. Only other investigations reported in relation to galactosaemia were liver function tests: peak bilirubin said to be higher and AST lower than other IMDs ($p<0.05$)	Median presenting age in galactosaemia: 7 days Presenting features: • n=17/17 jaundice • n=16/17 hepatomegaly • n=2/17 splenomegaly • n=3/17 hepatic encephalopathy Median age for all IMDs was 6 weeks (1 day to 41 months): • 7 days NPC • 14 days CDG type 1b • 1 month tyrosinemia type 1 • 5 months 3 weeks MRCD • 1 year 5 months OTC deficiency	No relevant information for KQ3. Reports at mean follow-up 4 years 3 months: • none of n=17 with galactosaemia received liver transplant • n=6/14 had no evidence of developmental disability* (43%) • n=8/14 had evidence of developmental disability* (57%) • n=3 had no follow-up data available *not further described <u>Q3 applicability</u> No further information available on treatment or disease outcomes or the relationship between the two.

Study reference Question	Study design	Population	Method of screen detection or diagnosis	Clinical presentation: age and symptoms	Outcomes and any relation to time of diagnosis/management
		GALT enzyme activity (unclear if any diagnoses previously known): • n=17 galactosaemia • n=7 mitochondrial respiratory chain disorder (MRCD) • n=4 ornithine transcarbamylase (OTC) deficiency • n=4 tyrosinemia type 1 • n=3 Niemann-Pick disease type C (NPC) • n=1 congenital disorder of glycosylation (CDG) type 1b			
Porta et al 2015[10] (NB detail in column 2 also applicable to Q2)	Retrospective NBS programme evaluation of cases identified with classic galactosaemia (CG) from Jan 1982 to Dec 2012 Italy, regional centre (Regional	n=1,123,909 screened n=8991 screen-positives recalled (0.80% recall rate) n=33 confirmed to have raised galactose at 2nd screen: • n=13 confirmed CG (mean GALT 2.1+/-8%) (incidence 1:86,000) • n=8 partial galactosaemia	<u>Screening</u> • 1982 to 2000: qualitative fluorometric assay • 2000 onwards: quantitative fluorescent galactose oxidase method • Cut-off: 10mg/dL total galactose (99th percentile based on analysis of 11,000 healthy newborns) Screen-positives recalled for second measure of total galactose.	NBS at 3.4 (+/-1.2) days of life Recall/2nd confirmation test at 12.2 (+/- 3.2) days (time of treatment initiation) <u>Clinical presentation</u> n=10/13 (76.9%) readmitted to hospital before availability of NBS results. Clinical symptoms occurred at mean 5.8 (+/- 1.1) days of life. Described as uniform with jaundice, lethargy, vomiting, and	Galactose restriction described to be implemented in all patients from time of 2nd confirmation (subsequently modified according to diagnosis) However the study describes that 'In all patients, diagnosis was formulated at our centre on the basis of clinical picture, further corroborated by availability of 1st tier screening results' - which suggests that management may have been commenced earlier.

Study reference Question	Study design	Population	Method of screen detection or diagnosis	Clinical presentation: age and symptoms	Outcomes and any relation to time of diagnosis/management
	Reference Center for Newborn Screening of Piemonte and Valle d'Aosta and the Regional Reference Center for diagnosis and treatment of IEM)	(GALT 17% to 49%; mean 27.0 +/- 1.6%) • n=7 transient galactosaemia (idiopathic) • n=3 with GALK deficiency • n=1 with GLUT2 (glucose transport 2) deficiency • n=1 with congenital porto-systemic shunt PPV=13/8991=0.14% for CG at 1st screen PPV= 13/33=39.4% for CG with confirmation at 2nd screen Mean galactose concentration at second screen was 102.4 (+/- 22.1) mg/dL in those with CG vs 29.3 (+/- 15.6) mg/dL in those with partial (p=0.005) Those with transient galactosaemia had galactose 29.3 (+/- 15.6) mg/dL at 2nd screen which gradually normalised after short	2nd confirmation: GALT enzyme activity assessment and clinical evaluation: • mean GALT 2.1% (+/- 8) • n=13 all had activity 0 to 6%; unclear if this was the upper cut-off for diagnosis NB. No reported genetic analysis. Diagnosis described to be based on combination of clinical symptoms corroborated by screening results.	liver failure.	Galactose-restriction was described to be 'effective in restoring the clinical picture over the next few days' <u>Long-term follow-up</u> Galactose described to steadily normalise to <3mg/dL in all patients n=10/13 (77%) described to have mild to severe intellectual disability at 15.2 (+/- 5.6) years n=3/13 (23%) with normal intellect (N=8/8 with partial galactosaemia all had 'adequate' intellect and development [at mean 9.2 years] with galactose 1.9 [+/-0.4] mg/dL) <u>Q3 applicability</u> No analysis of the relationship between time of treatment onset/diagnosis and outcomes. However, the study notes that 'Early acute decompensation was avoided in three siblings of affected children with severe GALT deficiency, put on galactose restriction immediately after DBS collection on the basis of presumptive diagnosis of classic galactosemia.' This was said to prevent galactose accumulation with earlier normalisation of galactose

Study reference Question	Study design	Population	Method of screen detection or diagnosis	Clinical presentation: age and symptoms	Outcomes and any relation to time of diagnosis/management
		period on galactose restriction			and an uncomplicated clinical course (mild jaundice only) in contrast with 'life-threatening' clinical course of the older child.
Ramadza et al 2018[8]	Retrospective case series of patients with CG identified in Croatia. Time period and centre unclear (Non-screening context)	n=16 patients with galactosaemia (8 male) with current age range 2.5 to 40 years.	All reported to have elevated galactose and severely reduced GALT activity without further specification. However, genotyping on whole blood samples has been performed for the full sample.	Clinical presentation at 2 days to 6 weeks: • n=8 at ≤7 days: n=3 at 2 days, n=1 each at 3, 4 and 5 days, n=2 at 7 days • n=3 at 1 to 2 weeks: at 9, 12 and 14 days • n=2 at 2 to 3 weeks: at 17 and 20 days • n=1 at 6 weeks • n=1 unspecified as 'during the neonatal period' • n=1 treated pre-emptively from birth due to an affected sibling Common symptoms across all ages of presentation were hyperbilirubinaemia, coagulopathy and other liver dysfunction and sepsis, sometimes with additional symptoms such as lethargy and cataracts Diagnosis at age range 7 days to 5 months:	Galactose-restricted diet said to be commenced immediately upon suspicion of diagnosis. n=7/10 aged >12 years reported to have complications, mainly neurological and developmental n=2/6 aged <12 years had long-term complications, n=1 speech delay and n=1 'severe mental retardation' (considered may be due to meningitis complications in the neonatal period) Of note n=1/16 managed pre-emptively from birth due to positive family history did not develop symptoms (homozygous for 2 pathogenic mutations p.Q188R). n=3 others had this genotype: n=2 presenting symptomatically at 2 days and n=1 at 20 days. Only the case presenting late at 20 days had chronic neurodevelopmental problems.

Study reference Question	Study design	Population	Method of screen detection or diagnosis	Clinical presentation: age and symptoms	Outcomes and any relation to time of diagnosis/management
				• n=1 at 7 days • n=6 at 1 to 2 weeks (n=3 day 8, day 10, 11 and 14) • n=4 at 3 to 4 weeks (day 21, 22, 26, 27) • n=3 at 6 weeks • n=1 at 5 months (Diagnostic time was not always consistent with the age of presentation, potentially because of the varied age range of cases and changes in diagnostic practice over time).	<u>Q3 applicability</u> Difficult to apply as evidence for the question of the relationship between time of treatment onset/diagnosis and outcomes. Galactose-restriction was said to be initiated on suspicion of diagnosis but it is unknown when this may have been, and there was also sometimes wide time interval between presenting symptoms and diagnosis (for example 2 with neonatal symptoms not diagnosed until 6 weeks and 1 at 5 months). Also the timing of presentation and later complications may bear more relation to genotype and variant severity rather than to treatment initiation.
Teke Kisa et al 2019[9]	Retrospective case series of patients with CG diagnosed and managed at 4 centres in Turkey, 1996 to 2017. (Non-screening context)	n=76 patients CG (41 male) with current age median 40 months (range 18 to 96 months) n=50 (86.2%) had homozygous genotypes and n=8 (13.8%) were compound heterozygotes. n=56 (74%) described to have parental consanguinity	All said to be diagnosed by: • GALT activity <10%, and/or • 2 pathogenic gene variants	Reports: • median age at first symptom: 10 days (25th to 75th centile 5 to 20 days) • median age at diagnosis: 30 days (25th to 75th centile 17 to 53 days; max 638 days) • n=40 (52.6%) diagnosed within 0 to 30 days • n=36 (47.4%) diagnosed >1 month • n=7 (9.2%) diagnosed at >3 months (presenting 3 days to 92 days, diagnosed 110 to 734 days)	All patients reported to have survived to follow-up (median 40 months). No analysis of relation between diagnostic time/start of treatment and follow-up symptoms. Reports the follow-up for the n=7 diagnosed at over 3 months of age: • Current age 3 to 11 years • their presenting symptom (mostly cataracts and jaundice) • reason for late diagnosis: n=3 no earlier symptoms, n=1 lost to follow-up, n=1 unclear, n=2 thought alternative diagnosis

Study reference Question	Study design	Population	Method of screen detection or diagnosis	Clinical presentation: age and symptoms	Outcomes and any relation to time of diagnosis/management
		(incidence in Turkey estimated in prior study as 1 in 23,775)		• n=16 described to be diagnosed within 2 weeks and started on diet • (unclear whether all were only started on diet at diagnosis or suspicion) Reports the following 'at presentation or time of diagnosis' • n=70 (92.0%) neonatal jaundice • n=32 (43.2%) neonatal sepsis • n=43 (66.2%) hepatomegaly • n=27 (54.2%) coagulopathy • n=24 (34.2%) cataract • (also gives median values for liver function and coagulation tests)	<u>Q3 applicability</u> Difficult to apply as evidence for the question of the relationship between time of treatment onset/diagnosis and outcomes: • outcomes are not given for those diagnosed earlier to compare with • most of the late presentations are due to later symptomatic presentation, while earlier presentation may be due to greater severity of disease (for example of genotype) • only n=3 were due to loss to follow-up/missed diagnosis: 1 with speech and developmental problems, 1 with speech problems and 1 with normal development • but in all cases actual treatment initiation is not described (for example whether some may have started presumptively before diagnosis) Notably genotype-phenotype correlation was described not to be seen.
Viggiano et al 2015[11]	Retrospective case series of patients with CG identified	n=19 patients with galactosaemia (11 male) with current age reported 1 to 34 years as defined by:	GALT activity <5% and confirmed genetic analysis gave diagnosis in all	n=13/14 were said to be 'symptomatic at the time of diagnosis': • n=13/13 with liver failure • n=8/13 with cataracts	n=6/7 aged >12 years reported to have complications including neurological, developmental and psychological

Study reference Question	Study design	Population	Method of screen detection or diagnosis	Clinical presentation: age and symptoms	Outcomes and any relation to time of diagnosis/management
	through NBS and referred to a single centre in Italy (University Hospital of Padua) from 1980 onwards.	• detection through NBS • diagnosis confirmed by absent/reduced GALT activity <5% • confirmed genetic analysis • galactose-free diet initiated upon disease suspicion • ongoing monitoring (though note as opposite only n=14 had CG as n=5 had Duarte variants)	Only n=14 had CG, all with 2 pathogenic alleles and GALT <1% n=5 carried Duarte alleles (1 homozygote, 2 compound with a normal allele, and 2 compound with a pathogenic allele (Detailed information is given on methods of DNA analysis, the genotypes identified and variants [18 variants] – not reported here as limited question applicability.)	• n=5/13 with sepsis • n=4/13 with anaemia Diagnosis is assumed to be genetic analysis/GALT level but the timing is not given. Also lack of clarity on NBS protocol and whether galactose restriction may have been initiated before diagnosis (for example if as Porta et al with 2nd confirmatory raised galactose)	n=1/7 aged <12 years had long-term complications though it was considered that symptoms may develop in the longer term even for the n=1/14 who had no symptoms at diagnosis. (All with Duarte were asymptomatic at diagnosis and no long-term complications reported). <u>Q3 applicability</u> No analysis of the relationship between time of treatment onset/diagnosis and outcomes.

Abbreviations: ALF, acute liver failure; AST, aspartate transaminase; CDG, congenital disorder of glycosylation; CG, classic galactosaemia; DBS, dried blood spot; Gal-1-P, galactose-1-phosphate; GALK, galactokinase; GALT, galactose-1-phosphate uridyltransferase; GLUT2, glucose transport 2; IEMs, inborn errors of metabolism; IMDs, inherited metabolic diseases; INR, international normalised ratio; MRCD, mitochondrial respiratory chain disorder; NBS, newborn screening; NPC, Niemann-Pick disease type C; OTC, ornithine transcarbamylase; PPV, positive predictive value; PT, prothrombin time

Table 16. Study applicable only to key question 2

Study reference Question	Study design	Screening method	Reference standard	Screened population	Test performance	Diagnosed cases/outcomes
Welling et al 2017[12]	Retrospective evaluation of the effectiveness of the NBS	DBS collected between 72 and 168 hours of birth and sent to one of	Diagnostic criteria: • Absent or barely-detectable residual red blood cell	Overall total screened: n=1,637,733 Screen positive: n=752	Overall test performance Sensitivity: 100%* Specificity: 99.56%	n=28 cases diagnosed through NBS plus n=3 pre-emptively diagnosed due to an affected sibling.

Study reference Question	Study design	Screening method	Reference standard	Screened population	Test performance	Diagnosed cases/outcomes
	programme for CG in the Netherlands between 2007 (inception) and 2015, with evaluation of the different screening methods used during this period. Secondary objective to look at the bio-clinical outcomes of those identified, including GALT level and genotypes.	5 regional laboratories. Screen positives sent to one of 7 Dutch centres. 5 methods over the study period as outlined below.	GALT activity (<15%); and/or • 2 pathogenic variants in the *GALT* gene Plus typical symptoms in the newborn period with persistently elevated Gal-1-P values despite dietary treatment. Some lack of clarity whether the latter points around (persistent) symptoms need to be present for diagnosis.	Diagnosed CG: n=28 (*plus 3) False positive: n=724 False negative: n=0 True negatives: n= 1,636,981 *n=3 with sibling history managed pre-emptively from birth.	Positive predictive value (PPV): 3.6% *NB. No cases of CG diagnosed clinically among screen-negatives during the study period; therefore assumed to be 100% sensitivity (negative predictive value not reported but assumed 100%)	Total n=31 (incidence 1:52,800) n=6 of these cases are classed as 'atypical' have previously unreported genotype and clinical characteristics Median age at referral for a positive screen was 6 days (range 3 to 10 days). Treatment started at median 6 days (range 0 to 10 days). Treatment was started immediately upon NBS in symptomatic cases but only after confirmation in asymptomatic infants. n=6/16 'typical' CG cases had neonatal cataract (0/6 atypical cases) which regressed in 3 to 14 months after dietary initiation Most typical CG cases had barely detectable GALT levels vs 3.6% to 9.3% in atypical cases. <u>Applicability to KQ1 or 3</u> KQ1: The study tabulates the n=31 cases and gives yes/no to 'symptoms of CG at diagnosis?' but it is unclear when diagnostic confirmation was achieved (only the median age at referral is given) and also unclear when the first symptoms occurred.
		Method 1 (Jan 2007 to April 2007): • Total galactose (galactose plus Gal-1-P) ≥700 µmol/l Total galactose measured using the Bio-Rad Quantase Neonatal Total Galactose screening assay.		Total screened: n=44,174 Screen positive: n=217 Diagnosed CG: n=1 False positive: n=216 False negative: n=0 True negatives: n=43,957	Sensitivity: 100% Specificity: 99.51% Positive predictive value (PPV): 0.46%	
		Method 2 (April 2007 to June 2012): • Tier 1: GALT ≤20%		Total screened: n=952,191 Screen positive: n=322 Diagnosed CG: n=18	Sensitivity: 100% Specificity: 99.97% PPV: 5.6%	

Study reference Question	Study design	Screening method	Reference standard	Screened population	Test performance	Diagnosed cases/outcomes
		• Tier 2: total galactose ≥700 µmol/l GALT measured using the using the Bio-Rad CPDA Neonatal GALT assay. Patients referred when both abnormal (assume both was criteria for screen positives, rather than one or other)		False positive: n=304 False negative: n=0 True negatives: n=951,869		KQ3: no analysis of timing of treatment initiation in relation to symptoms.
		Method 3 (July 2012 to June 2014): • Tier 1: GALT ≤15% • Tier 2: total galactose ≥700 µmol/l In 2012/13, three labs switched to the automated GALT assay using the Genetic Screening Processor (GSP) and two switched to the manual		Total screened: n=345,685 Screen positive: n=87 Diagnosed CG: n=6 False positive: n=81 False negative: n=0* *n=1 diagnosed prior to birth with galactose-free diet from day 1 (total galactose normal range) True negatives: n=345,598	Sensitivity: 100% Specificity: 99.98% PPV: 6.9%	

Study reference Question	Study design	Screening method	Reference standard	Screened population	Test performance	Diagnosed cases/outcomes
		GSP assay (due to technical issues with the Biorad GALT assay) (said to have good correlation).				
		Method 4 (July 2014 to June 2015): • Tier 1: GALT ≤2.7 units/decilitre • Tier 2: total galactose ≥900 µmol/l (same assays as method 3)		Total screened: n=173,656 Screen positive: n=96 Diagnosed CG: n=3 False positive: n=93 False negative: n=0* *n=2 diagnosed prior to birth with galactose-free diet from day 1 (total galactose normal range) True negatives: n=173,560	Sensitivity: 100% Specificity: 99.95% PPV: 3.1%	
		Method 5 (July 2015 to Dec 2015): • Tier 1: GALT ≤2.0 U/dl • Tier 2: total galactose ≥1000 µmol/l (same assays as method 3)		Total screened: n=122,027 Screen positive: n=30 Diagnosed CG: n=0 False positive: n=30 False negative: n=0 True negatives: n=121,997	Sensitivity: Unknown Specificity: 0%* (as reported) PPV: Unknown *Sensitivity and PPV are unknown because there were no cases. Apparent error in reporting specificity, which is 100% (121,997/122,027)	

Abbreviations: CG, classic galactosaemia; DBS, dried blood spot; Gal-1-P, galactose-1-phosphate; GALT, galactose-1-phosphate uridyltransferase; GPS, genetic screening processor; NBS, newborn screening; PPV, positive predictive value

Table 17. Studies relevant to key question 3

Study reference Question	Study design	Population	Method of screen detection or diagnosis	Clinical presentation: age and symptoms	Outcomes and any relation to time of diagnosis/management
Lak et al 2017[13] KQ3	Cochrane systematic review to assess whether there is evidence that NBS for galactosaemia prevents or reduces mortality and morbidity and improves clinical outcomes in affected neonates, and the quality of life in older children.	Eligible studies: RCTs and quasi-RCTs where participants were prospectively allocated to either screening via a blood test (heel-prick or venous sample) or to no screening. Eligible population: All newborns eligible for screening in the first week of life. Eligible tests: any blood test to measure total galactose, GAL-1-P [GALT measure not specified] including fluorescent spot test (Beutler), calorimetric, fluorescent galactose oxidase method, Guthrie's method, etc. Excluded: urine and genetic testing. No language restrictions, published	No eligible studies identified	No eligible studies identified	No eligible studies identified Planned outcomes were: Primary: • mortality (disease-related) • neonatal morbidity (liver failure or sepsis) Secondary outcomes: • quality of life • clinical outcomes (organ dysfunction, neurodevelopmental) • reduction in Gal-1-P levels

Study reference Question	Study design	Population	Method of screen detection or diagnosis	Clinical presentation: age and symptoms	Outcomes and any relation to time of diagnosis/management
		and unpublished literature. Search to 18th December 2017 of CENTRAL and PubMed Additionally: • US National Institutes of Health Ongoing Trials Register Clinicaltrials.gov • World Health Organization International Clinical Trials Registry Platform • Grey Literature Report System for Information on Grey Literature in Europe (plus planned hand-searching of retrieved literature)			
Ozgun et al 2019[14]	Retrospective record review/case series of CG patients homozygous for p.Gln188Arg variant seen in	n=68 patients apparently with diagnosis of CG but not specific to variant n=46 homozygous for p.Gln188Arg who had detailed neurological assessment, MRI and	'GALT activity <3U/g haemoglobin and gene analysis upon clinical suspicion' Lactose-free diet (with restriction of other sources) used, though does not specify	No detail applicable to KQ1; age at presentation and symptoms are not specified (only age at diagnosis)	<u>Developmental</u> Denver II developmental screening test for children aged 0 to 6 years. Measured: • Personal–Social: communication and self-care abilities

Study reference Question	Study design	Population	Method of screen detection or diagnosis	Clinical presentation: age and symptoms	Outcomes and any relation to time of diagnosis/management

	a single Turkish hospital, 2003 to 2017. Specific number of centres unclear.
	
	Non-screening context
		developmental assessment and no other disease were included.
		
		n=20 diagnosed ≤1 week (1 to 7 days)
		
		n=26 diagnosed >1 week:
		• n=10 in 8 to 14 days
		• n=7 in 15 to 21 days
		• n=4 in 22 to 28 days
		• n=2 in 29 to 35 days
		• n=3 in 36 to 42 days
		
		n=5 diagnosed because of an affected sibling; no other detail on presentation
			whether this was on diagnostic confirmation or suspicion.		• Fine Motor: Hand-eye coordination, manipulation of objects, problem solving
					• Language: Receptive and expressive language and hearing
					• Gross Motor: Body movements, such as sitting, walking, overall coordination.
					
					Developmental ages were determined as the age of the latest 'pass' item before two consecutive 'fail' items.
					
					Developmental delay diagnosed if developmental age is ≥20% lower than chronological age
					
					Mean age at latest test 34.4 +/- 18.2 months:
					• age-appropriate development: n=25 (54.3%)
					• delay in at least 1 domain: n=21 (45.7%):
					  ○ most common language (n=19, 41.3%)
					  ○ speech only (n=9, 19.5%)
					
					Time of diagnosis in those with delay: 2.38 ± 1.75 weeks vs 2.20 ± 1.35 weeks for those with normal examination (p=0.954)

Study reference Question	Study design	Population	Method of screen detection or diagnosis	Clinical presentation: age and symptoms	Outcomes and any relation to time of diagnosis/management
					(no association between developmental delay and gender or abnormality on MRI) <u>Neuroimaging by MRI scan</u> Mean age of patients having MRI was 37.7 +/- 17.7 months Mean age of most recent MRI was 48.5 +/- 28.5 months. Most frequent finding was a change in signal intensity in the periventricular white matter (n=8 36.3%) followed by ventricular enlargement (n=7, 31.8%) and thinning corpus callosum (n=5, 22.7%) (associated with developmental delay p≤0.05) (age at diagnosis not specified for normal vs abnormal) <u>Neurological examination</u> n=11/46 (23.9%) abnormal: • n=6 ataxia • n=2 tremor • n=1 chorea and dystonia • n=1 hypotonia • n=1 horizontal nystagmus Time of diagnosis 2.9 ± 1.7 weeks vs 2.0 ± 1.4 weeks for those with normal examination (p=0.137)

Study reference Question	Study design	Population	Method of screen detection or diagnosis	Clinical presentation: age and symptoms	Outcomes and any relation to time of diagnosis/management
					(no association with MRI pathology or developmental delay) <u>Timing of diagnosis</u> There was no correlation between the time of diagnosis and the presence of pathology in neurological examination (p=0.137, r=0.031), developmental delay (p=0.954, r= -0.028) and presence of MRI abnormality (p=0.917, r= -0.059) When examining by group diagnosed within or beyond 1 week there was no difference in neurological complications: • Number with developmental delay: n=10/20 diagnosed within 1 week vs n=11/26 beyond; p=0.796 • Pathology on brain MRI: n=10/20 vs n=12/26; p=0.515 • Abnormal neurological exam (no detail): n=3/20 vs n=8/26; p=0.187 • Convulsion (no detail): n=6/20 vs n=5/26; p=0.307 Mann-Whitney U test for binary comparisons; Wilcoxon rank-sum for dependent groups; Spearman's test for correlation analysis. Statistical significance set at p<0.05

Study reference Question	Study design	Population	Method of screen detection or diagnosis	Clinical presentation: age and symptoms	Outcomes and any relation to time of diagnosis/management
Rubio-Gozalbo et al 2019[15]	Evaluation of those with CG recorded in the International Galactosaemia Networks (GalNet) Registry. International coverage of 15 countries (32 centres): Austria, Belgium, Croatia, Estonia, France, Germany, Ireland, UK, Israel, Lithuania, Netherlands, Portugal, Switzerland, Spain, US. Established 2014, coverage to 2019.	n=509 patients (48.1% male) • median age 18.0 years (range 0 to 65 years): 45.8% aged <18 years and 54.2% aged ≥18 years • 93.6% Caucasian • no information on sibling relationships	Including only patients with: • GALT activity ≤10%; and/or • pathogenic variants (1 or 2) Reported GALT activity: • n=211 (82.7%) ≤1% • n=36 (14.1%) >1 to ≤5% • n=8 (3.1%) >5 to ≤10% Genotypes listed (most common homozygosity p.Gln188Arg; 57.7%). n=215/468 (45.6%) reported to be diagnosed following NBS but methods not reported and likely to vary. n=89/98 (90.8%) reported to have elevated neonatal Gal-1-P (>10mg/dL haemoglobin) No further breakdown of subjects by screened vs non-screened	Reports n=332/416 (79.8%) had acute neonatal illness • defined as ≥1 of: encephalopathy, bleeding diathesis, signs of infection, elevated liver enzymes or hypoglycaemia (81.7% of n=509 sample) Specifics listed as 'neonatal illness': • n=211/300 (70.3%) elevated liver enzymes • n=128/301 (42.5%) bleeding diathesis (abnormal clotting) • n=71/245 (29.0%) encephalopathy • n=96/351 (27.4%) infection (of which 56.3% of those assessed had positive blood culture) • n=68/264 (25.8%) cataract • n=65/259 (25.1%) hypoglycaemia <u>No information on age at presentation to give applicability to KQ1</u> Most infants given soy (n=302/394, 76.6%), 12.7% (n=50/394) given elemental and 10.7% (n=50/394) other galactose-restricted formula.	<u>Q3 applicability</u> Reports the following were associated with lower risk of neonatal complications: • diagnosis following NBS: odds ratio (OR) 0.30, 95% confidence interval (CI) 0.20 to 0.47 (p<0.0000001) 215/468 (45.9%) vs 253/468 (54.1%) non-screened (92% with info on screening status available, though not all had data on neonatal complications) • initiation of galactose restriction within the first week of life: OR 0.32, 95% CI 0.21 to 0.50 (p<0.000001) 198/391 (50.6%) within 1 week vs 193/391 (49.4%) >1 week (77% with info on time of dietary initiation, though not all had data on neonatal complications) Patients diagnosed following NBS were said to be 'often younger (p<0.000001) and started diet in the first week of life (p<0.000000000001).' Discussion also described 'neurological complications were less prevalent in patients diagnosed following NBS': OR 0.32, 95% CI 0.20 to 0.51 (p<0.00001)

Study reference Question	Study design	Population	Method of screen detection or diagnosis	Clinical presentation: age and symptoms	Outcomes and any relation to time of diagnosis/management
				Diet was implemented on the first day for 16.6% (n=65/391), in the first week for 33.9% (133/39), in the second week for 34.2% (134/391), in the 3rd to 4th week for 9.4% (37/391) and after 28 days for 5.9% (23/391). Most patients followed a lactose-free diet, ongoing (94.2%, n=406/431). The majority (64.3%, n=245/381) adhered to a relaxed diet (lactose-free without further restrictions), rather than a strict diet (lactose-free and restriction of non-dairy sources) which was followed by 35.7%.	(only 63.4% had neurological assessment) The method of statistical analysis is described: Fisher's exact test for categorical variables with all outcomes said to be grouped into 2 categorical groups (present vs absent) Also describes methods to handle the missing observations including performing available case analysis (for descriptive analysis) and complete case analysis (for association analysis, odds ratios and Fisher's exact test). However, absolute numbers are not available within this publication (including appendices) on characteristics and long-term outcomes of screened vs non-screened. All characteristics are given for the registry group as a whole. Frequency of complications for group as a whole includes: • n=167/320 (52.2%) developmental delay • n=192/289 (66.4%) language and speech disorders

Study reference Question	Study design	Population	Method of screen detection or diagnosis	Clinical presentation: age and symptoms	Outcomes and any relation to time of diagnosis/management
					• n=168/323 (52.0%) neurological complications • n=128/288 (44.4%) psychiatric and behavioural problems Gonadal complications: Female • n=65/134 induced puberty (48.5%) (remainder spontaneous) • n=118/148 (79.7%) primary ovarian insufficiency Male • n=3/63 (4.8%) delayed puberty • n=3/54 (5.6%) cryptorchidism Low bone mineral density (BMD T score ≤ -1.0 standard deviation or BMD Z score ≤ -2.0 standard deviations) : • n=76/287 (26.5%) (65.8% female) Cataract was reported for n=68 during the neonatal period. Residual cataract at later follow-up was reported for n=20/44 (45.5%) with new development reported for n=22/238 (9.2%).

Abbreviations: BMD, bone mineral density; CG, classic galactosaemia; CI, confidence interval; DBS, dried blood spot; Gal-1-P, galactose-1-phosphate; GALT, galactose-1-phosphate uridyltransferase; GalNet, international galactosaemia networks; MRI, magnetic resonance imaging; NBS, newborn screening; OR, odds ration; RCTs, randomised controlled trial

Appraisal for quality and risk of bias

Quality assessment of case series was based on the adaption of the Newcastle Ottawa scale for cohort and case-control studies (with removal of items related to comparability and adjustment), Bradford Hills and Pierson criteria, as described by Murad et al.[22] Notably the overall judgement on quality and applicability was primarily based on relation to the key question, and does not necessarily represent the quality of the study for its overall purpose.

Studies providing test performance data were also assessed using QUADAS-2, though these studies were retrospective programme evaluations/case series and not typical prospective diagnostic accuracy studies. Therefore, again, any quality limitations in regard to these assessments do not necessarily reflect limitations of the study for its overall purpose.

Table 18. Quality assessment of Hegarty et al[7] case series for KQ1

Domain	Question	Judgement on risk of bias	Comment
Selection	1. Do the patients represent the whole experience of the investigator (centre) or is the selection method unclear to the extent that other patients with similar presentation may not have been reported?	Low	Acute liver failure (ALF) is clearly defined so it can be assumed that this study represents all children <5 years who presented with ALF during this time-period – with the caveat that this only represents children presenting with ALF
Ascertainment	2. Was the exposure* adequately ascertained?	Unclear	*The 'exposure' here considered to be classic galactosaemia (CG): the study reports measuring GALT activity with no further information on cut-off or overall diagnostic criteria
	3. Was the outcome adequately ascertained?	Unclear	Age at presentation: 7 days is reported as the median but no further information is given Long-term outcomes: limited reporting on clinical features but does not affect applicability to KQ1
Causality	4. Were other alternative causes that may explain the observation ruled out?	NA	Questions 4, 5 and 6 mostly applicable to reporting of drug adverse effects
	5. Was there a challenge/re-challenge phenomenon?	NA	
	6. Was there a dose–response effect?	NA	

	7. Was follow-up long enough for outcomes to occur?	Low	For presentation, the study included all children <5 years presenting with ALF which would be sufficient time for CG to present Follow-up was to 4 years for longer-term outcomes but it is not possible to relate to treatment.
Reporting	8. Are the cases described with sufficient details to allow other investigators to replicate the research or to allow practitioners make inferences related to their own practice?	High	Insufficient detail is given on CG diagnostic criteria, management or follow-up outcomes. (As above the study also has limited applicability in that it only represents cases of CG presenting with ALF)
Overall judgement on methodological quality and applicability to question: fair/poor – primarily in relation to lack of definition of CG diagnostic criteria and the narrow representation of all children with CG			

Table 19. Quality assessment of Porta et al[10] case series for KQ1

Domain	Question	Judgement on risk of bias	Comment
Selection	1. Do the patients represent the whole experience of the investigator (centre) or is the selection method unclear to the extent that other patients with similar presentation may not have been reported?	Low	No indication that the evaluation is selective and it appears to cover all cases diagnosed in this region during the 30 year period.
Ascertainment	2. Was the exposure* adequately ascertained?	Low	*The 'exposure' here considered as CG: The study does not report genetic analysis which is reported by criteria to be required for diagnostic confirmation. However, the basis of 2 x raised galactose, deficient GALT combined with compatible clinical symptoms appears sufficient for diagnosis.
	3. Was the outcome adequately ascertained?	Low	Age at screening, confirmation and reporting is clearly reported. Long-term outcomes: limited reporting on clinical features but does not affect applicability to KQ1 or 2
Causality	4. Were other alternative causes that may explain the observation ruled out?	NA	Questions 4, 5 and 6 mostly applicable to reporting of drug adverse effects

	5. Was there a challenge/re-challenge phenomenon?	NA	
	6. Was there a dose–response effect?	NA	
	7. Was follow-up long enough for outcomes to occur?	Low	Follow-up to young adulthood
Reporting	8. Are the cases described with sufficient details to allow other investigators to replicate the research or to allow practitioners make inferences related to their own practice?	Low	Sufficient detail is given on CG diagnostic criteria and summary of management and follow-up outcomes.
Overall judgement on methodological quality and applicability to question: good			

Table 20. Quality assessment of Ramadza et al[8] case series for KQ1

Domain	Question	Judgement on risk of bias	Comment
Selection	1. Do the patients represent the whole experience of the investigator (centre) or is the selection method unclear to the extent that other patients with similar presentation may not have been reported?	Unclear	The centre, time period and selection of patients are not reported. Therefore it is unclear how representative this is.
Ascertainment	2. Was the exposure* adequately ascertained?	Low	*The 'exposure' here considered as CG: There is no explanation of the diagnostic process at the time of presentation or diagnosis in terms of total galactose or GALT levels. However, genotypes have been analysed and are given so the carriage of 2 pathogenic alleles would give the diagnosis according to current criteria.
	3. Was the outcome adequately ascertained?	Low	Age at clinical presentation and diagnosis is given. Long-term outcomes: adequate reporting on clinical outcomes at current age but it is not possible to easily relate this to time of treatment initiation as this is not specified or analysed.
Causality	4. Were other alternative causes that may explain the observation ruled out?	NA	Questions 4, 5 and 6 mostly applicable to reporting of drug adverse effects
	5. Was there a challenge/re-challenge phenomenon?	NA	
	6. Was there a dose–response effect?	NA	

	7. Was follow-up long enough for outcomes to occur?	Low	Range of ages at follow-up to young adulthood
Reporting	8. Are the cases described with sufficient details to allow other investigators to replicate the research or to allow practitioners make inferences related to their own practice?	Unclear	Some lack of clarity around the presentation, and reasons to delay in diagnosis, or whether this case series is fully representative.
Overall judgement on methodological quality and applicability to question: fair as age at presentation and diagnosis are clearly given for each patient but there is some lack of detail (such as GALT and galactose levels and uncertainty over the representation).			

Table 21. Quality assessment of Teke Kisa et al[9] case series for KQ1

Domain	Question	Judgement on risk of bias	Comment
Selection	1. Do the patients represent the whole experience of the investigator (centre) or is the selection method unclear to the extent that other patients with similar presentation may not have been reported?	Low	No indication that the evaluation is selective and it appears to cover all cases diagnosed in these 4 centres during the study period.
Ascertainment	2. Was the exposure* adequately ascertained?	Low	*The 'exposure' here considered as CG: Diagnostic criteria are compatible with the current gold standard.
	3. Was the outcome adequately ascertained?	Low	Age at presentation and diagnosis is clearly reported. Long-term outcomes: lack of clarity around treatment initiation and difficulty in applying the evidence for KQ3 but does not affect applicability to KQ1
Causality	4. Were other alternative causes that may explain the observation ruled out?	NA	Questions 4, 5 and 6 mostly applicable to reporting of drug adverse effects
	5. Was there a challenge/re-challenge phenomenon?	NA	
	6. Was there a dose–response effect?	NA	
	7. Was follow-up long enough for outcomes to occur?	Low	There is an age range of cases but including follow-up of most cases into childhood and adulthood
Reporting	8. Are the cases described with sufficient details to allow other investigators to replicate the research or to allow practitioners make inferences related to their own practice?	Unclear	Sufficient detail is given on CG diagnostic criteria and summary of presentation and follow-up outcomes. However, this may not fully reflect the non-screening context in the UK as the incidence is higher (estimated 1

			in 23,775 due to high parental consanguinity)
Overall judgement on methodological quality and applicability to question: good quality study but some applicability issues			

Table 22. Quality assessment of Viggiano et al[11] case series for KQ1

Domain	Question	Judgement on risk of bias	Comment
Selection	1. Do the patients represent the whole experience of the investigator (centre) or is the selection method unclear to the extent that other patients with similar presentation may not have been reported?	Low	No indication that the evaluation is selective and it appears to cover all cases referred to this centre meeting the specified criteria
Ascertainment	2. Was the exposure* adequately ascertained?	Low	*The 'exposure' here considered as CG: Based upon profound GALT deficiency and 2 pathogenic variants
	3. Was the outcome adequately ascertained?	High	For the purposes of this analysis, age at diagnosis and how it relates to the stage of the NBS process is unclear. Long-term outcomes: adequate reporting on clinical features but this cannot be related to time of treatment initiation to provide evidence for Q3.
Causality	4. Were other alternative causes that may explain the observation ruled out?	NA	Questions 4, 5 and 6 mostly applicable to reporting of drug adverse effects
	5. Was there a challenge/re-challenge phenomenon?	NA	
	6. Was there a dose–response effect?	NA	
	7. Was follow-up long enough for outcomes to occur?	Low	Range of ages at follow-up to young adulthood
Reporting	8. Are the cases described with sufficient details to allow other investigators to replicate the research or to allow practitioners make inferences related to their own practice?	Unclear	Sufficient detail is given on CG diagnostic criteria and summary of management and follow-up outcomes, but there is lack of clarity around timing of presentation and treatment initiation.
Overall judgement on methodological quality and applicability to question: fair/poor as age at presentation is not given and specific relation in timing of NBS process is unclear (initial stages of which are not clarified)			

Table 23. QUADAS-2 assessment of Porta et al[10] test performance data for KQ2

Domain	Signalling question	Signalling question: Yes, no, unclear, not applicable Bias – low, high, unclear	Comment
Selection	Was a consecutive or random sample of the population enrolled?	Yes	
	Was a case-control design avoided?	Yes	
	Inappropriate exclusions avoided?	Yes*	*Not specified but no indication that this has excluded any cases of CG of any severity
Risk of bias	Could patient selection have introduced bias?	Low risk	
Applicability concerns	Is there concern that the included participants do not match the review question?	Low risk	
Index test	Index test results interpreted without knowledge of reference standard results?	Yes	
	Threshold pre-specified?	Yes*	*Though note unclear how this would have been formed initially as based on the 99th centile for this first 11,000 healthy newborns
Risk of bias	Could the conduct or interpretation of the index test have introduced bias?	Low risk	Assumed based on available content within publication but as above it is not possible to understand initial screening practice upon implementation
Applicability concerns	Is there concern that the index test, its conduct or interpretation differ from the review question?	Unclear risk	A range of tests, assays and cut-offs could be used in other populations.
Reference standard	Reference standard likely to correctly classify condition?	Unclear	The diagnosis was based on 2 x raised galactose, deficient GALT combined with compatible clinical symptoms. This appears sufficient for diagnosis, though the threshold for GALT deficiency was not clearly given nor genetic analysis reported. It is also unclear whether there was any threshold in symptomatic level. Therefore overall not possible to say with certainty.
	Reference standard results interpreted without knowledge of index test results?	No	Previous raised galactose on 2 occasions was indication for GALT enzyme assessment
Risk of bias	Could the conduct or interpretation of the reference standard have introduced bias?	Unclear risk	
Applicability concerns	Is there concern that the target condition as defined by the reference standard does not match the review question?	Unclear risk	Expected to give accurate diagnosis based on galactose, GALT and symptoms, though not specifically in accordance with

			recommended definitive diagnosis so difficult to say whether other reference standard may be used elsewhere.
Test strategy flow and timing	Was there an appropriate interval between the index test and reference standard?	NA	The condition could be diagnosed at the time of the index test when there would have been galactose exposure, and also at later follow-up. Therefore this is not expected to have an effect.
	Did all participants receive the same reference standard?	No	Those who screened negative at first screen and 2nd recall screen were not further followed up. There are not expected to have been screen negatives within this group as it is expected galactose would be raised. However, as this was an analysis of the NBS programme, without the study specifying, it is not possible to say that this evaluation would have included any clinically diagnosed (even though this may be unlikely).
	Were all participants included in analysis?	No	As above, the analysis does not include screen negatives at first screen or 2nd recall screen.
Risk of bias	Could the participant flow have introduced bias?	Unclear risk	Overall although false negative are expected unlikely, it is not possible to assume 100% sensitivity.
Overall judgement on methodological quality and applicability to question: low risk around selection and index test, but overall unclear risk around reference standard, participant flow and applicability.			

Table 24. QUADAS-2 assessment of Welling et al[12] for KQ2

Domain	Signalling question	Signalling question: Yes, no, unclear, not applicable Bias – low, high, unclear	Comment
Selection	Was a consecutive or random sample of the population enrolled?	Yes	
	Was a case-control design avoided?	Yes	
	Inappropriate exclusions avoided?	Yes	
Risk of bias	Could patient selection have introduced bias?	Low risk	
Applicability concerns	Is there concern that the included participants do not match the review question?	Low risk	

Index test	Index test results interpreted without knowledge of reference standard results?	Yes	
	Threshold pre-specified?	Yes*	*Though 5 different methods were used.
Risk of bias	Could the conduct or interpretation of the index test have introduced bias?	Low risk	No apparent indication of bias – though the test/threshold were continually modified
Applicability concerns	Is there concern that the index test, its conduct or interpretation differ from the review question?	Unclear risk	A range of tests, assays and cut-offs could be used in other populations.
Reference standard	Reference standard likely to correctly classify condition?	Yes	The diagnosis was based on recognised criteria of profoundly deficient GALT activity combined with 2 pathogenic variants. There is though some lack of clarity in that it also says plus newborn symptoms and elevated Gal-1-P values despite dietary treatment, making it unclear whether persistent symptoms needed to be present.
	Reference standard results interpreted without knowledge of index test results?	No	Previous raised galactose with/without raised GALT activity was the indication for referral.
Risk of bias	Could the conduct or interpretation of the reference standard have introduced bias?	Low risk	Despite some lack of clarity, overall there is considered unlikely to be much risk of bias in establishing the diagnosis given the use of the recognised criteria.
Applicability concerns	Is there concern that the target condition as defined by the reference standard does not match the review question?	Low risk	The study has used recognised diagnostic criteria (given some uncertainty about whether symptoms also needed to be present, as above).
Test strategy flow and timing	Was there an appropriate interval between the index test and reference standard?	NA	The condition could be diagnosed at the time of the index test when there would have been galactose exposure, and also at later follow-up. Therefore this is not expected to have an effect.
	Did all participants receive the same reference standard?	Yes	
	Were all participants included in analysis?	No	Screen negatives have not received the reference standard. However, the study has reported no cases were clinically diagnosed during the follow-up period.
	Could the participant flow have introduced bias?	Low risk	

Overall judgement on methodological quality and applicability to question: low overall risk of bias, but some uncertainty around applicability due to the range of index tests/thresholds that were used that were continually updated over the screening period.

Table 25. CASP quality assessment of Lak et al[13] systematic review for KQ3

Domain	Question	Yes, no, unclear, not applicable	Comment
Study validity	1. Did the review address a clearly focused issue?	Yes	
	2. Did the authors look for the right type of papers?	Yes	
	3. Do you think all the relevant studies were included?	Yes	
	4. Did the review's authors do enough to assess quality of the included studies?	NA	No studies identified
	5. If the results of the review have been combined, was it reasonable to do so?	NA	
The results	6. What are the overall results of the review?	NA	
	7. How precise are the results?	NA	
Applicability	8. Can the results be applied to the local population?	NA	
	9. Were all important outcomes considered?	NA	
	10. Are the benefits worth the harms and costs?	NA	
Overall judgement on methodological quality and applicability to question: high quality review but no eligible literature			

Table 26. CASP quality assessment of Ozgun et al[14] cohort study* for KQ3

*This case series has been assessed using a comparative cohort quality assessment tool because the study provides evidence for the question of whether early treatment improves outcomes compared with late treatment. This assessment does not diminish the value of the study for its primary purpose as a descriptive case series.

Domain	Question	Yes, no, unclear, not applicable	Comment
Study validity	1. Did the study address a clearly focused issue?	No	The study was designed to examine the early developmental status and neurological findings for children with CG homozygous for this variant. It was not a prospective study designed with the specific purpose of assessing whether time of diagnosis was associated outcome.
	2. Was the cohort recruited in an acceptable way?	Unclear	It reports to have reviewed hospital records at this single centre 2003 to 2017 to identify children homozygous for this variant. N=68

			were identified (presumably all CG though unclear whether all with this variant) and n=46 had this variant and the applicable data. It is expected to be representative and all would have received these assessments but it is unclear.
	3. Was the exposure adequately measured to minimise bias?	Unclear	The diagnosis of CG is compatible with the gold standard. The report evaluates 'time of diagnosis.' The time of treatment is not completely clear. For example whether galactose-restriction may have been introduced only on diagnostic confirmation through DNA analysis or at initial clinical suspicion on presentation, and how much delay there may have been between the two.
	4. Was the outcome accurately measured to minimise bias?	Yes	Those with data available all appear to have received this battery of tests with outcomes clearly specified.
	5a. Have the authors identified important sources of confounding?	Yes*	*The study looked at whether individual factors such as gender, age at last test, and age at diagnosis, and other test results were associated with each assessment. Dietary practice was described the same for all, all had the same variant and were treated at the same centre. So although not looking for confounding influence of other factors you may expect this to be a fairly homozygous group.
	5b. Have the authors taken account of the confounding factors in the designs and/or analysis?	No*	*In looking at the association with time of diagnosis other factors have not been accounted for, though as above aside from current age and gender most should be standardised. Current age was within a narrow range.
	6a. Was the follow up of subjects complete enough?	Unclear	It is unclear how many with this genotype may not have had all tests performed and whether there may be differences between those with and without comprehensive examination.

	6b. Was the follow-up of subjects long-enough?	Yes*	*Children were 3 to 4 years old at the time of last test. Therefore these are only outcomes to young childhood but this is a current sample and not likely to affect the outcome association.
The results	7. What are the results?	Unclear	The diagnostic age of those with and without developmental and neurological problems, and the proportion of those with problems who were diagnosed before and after 1 week, are given. P values are given for the differences but it has not been possible to give relative risk associations.
	8. How precise are the results?	Unclear	As above was no significant association but these are small samples and risk associations cannot be given.
	9. Do you believe the results?	Unclear	As above.
Applicability	10. Can the results be applied to the local population?	Unclear	This is applicable to children homozygous for this genotype in this centre in Turkey. The results may be applicable to others with the same genotype in Turkey, but due to care differences they could not be automatically be applied to other populations. Nor could they be applied to those with other variants.
	11. Do the results of this study fit with other available evidence?	No	Some prior evidence has been mixed though there is a lack of recent high-quality prospective comparative studies.
	12. What are the implications of this study for practice?	Unclear	Due to the various caveats above.
Overall judgement on methodological quality and applicability to question: overall low quality and applicability. It is a relatively homozygous sample with minimal expected influence from other confounders. However, there is potential selection bias as this only represents those who were tested, the sample is small decreasing the reliability of analyses and there is uncertainty over the actual time of dietary initiation and whether this may have differed from time of diagnosis. Also uncertain applicability to other countries.			

Table 27. CASP quality assessment of Rubio-Gozalbo et al[15] cohort study* for KQ3

*This is primarily a registry study/non-controlled cohort. It has been assessed using a comparative cohort quality assessment tool because the study provides evidence for the question of whether early treatment improves outcomes compared with late treatment. This assessment does not diminish from the value of the study as a whole as a retrospective evaluation of the characteristics of cases within the GalNet registry.

Domain	Question	Yes, no, unclear, not applicable	Comment
Study validity	1. Did the study address a clearly focused issue?	No	This was not a prospective cohort study designed with the purpose of assessing whether certain characteristics or exposures are associated with different outcome. It was purely aiming to review the natural history of CG using a large dataset of patients.
	2. Was the cohort recruited in an acceptable way?	Unclear	The study covers 15 countries (mostly European, US and Israel) and 32 centres within these countries. Without explicit mention in the report, it is not possible to know how complete this registry is and the estimated coverage in terms of all eligible European/western populations. For example, some countries are not covered (such as Italy, Canada, Scandinavia and Australasia), while for included countries, it is unknown whether this covers all eligible CG centres (for example only one US centre is included).
	3. Was the exposure adequately measured to minimise bias?	Unclear	Diagnosis of individuals is not covered and likely to vary, diagnostic criteria for CG are compatible with the current gold standard. When defining the 'exposure' as whether or not NBS was performed and time of dietary initiation, these are from medical records and expected to be reliable – however, information on NBS was only available for 92% and time of diet for 77% of the cohort.
	4. Was the outcome accurately measured to minimise bias?	Unclear	As a retrospective registry study, the authors themselves note caution that not all patients have been followed in a standardised, systematic manner, nothing that not all had received neurological examination by a neurologist. Valid outcome data is available for only a proportion of the registry sample (82% for neonatal complications and for neurological

			63%) – and within those numbers it is unclear how many had exposure data available.
	5a. Have the authors identified important sources of confounding?	No	It is not reported whether the associations with NBS and timing of dietary initiation were additionally adjusted for factors such as patient age (birth cohort), country, enzyme activity or strictness of diet.
	5b. Have the authors taken account of the confounding factors in the designs and/or analysis?	No	As above, none are reported.
	6a. Was the follow up of subjects complete enough?	Unclear	As point 4, not all patients may have been followed up in the same way. It is also unknown whether there may have been more comprehensive reporting of individuals with complications than those without.
	6b. Was the follow-up of subjects long-enough?	Yes*	*This represented an age range 0 to 65 years (median 18.0 years) with 45.8% aged under 18 and 54.2% aged over 18 years. Therefore not all will have long term data available, but this is expected to typical for what could be obtained from a registry sample.
The results	7. What are the results?	Unclear	Although the number screened and implementing treatment within the first week of life is known (roughly half the total sample in both cases), the absolute numbers with complications among screened vs non-screened or early vs late treatment is not given. Absolute rate reductions are unclear.
	13. How precise are the results?	Unclear	There seems to be a clear and strong association but difficult to assess as low risk given the uncertainty around numbers
	14. Do you believe the results?	Unclear	It is a large effect, but due to the uncertainty around whether follow-up is complete/representative, lack of clarity around absolute numbers and uncertain potential for confounding, it is difficult to know with certainty whether the outcome

			can be attributed to the difference in exposure. There is also no assessment of the association with other long term outcomes aside from neurological.
Applicability	15. Can the results be applied to the local population?	Unclear	There is likely variation in the screening methods used across the countries, and uncertainty whether the registry is fully comprehensive, whether outcomes are complete and whether there could be other influencing factors, leaving outstanding questions. Ideally a study may need to compare screened vs non-screened to see whether there is a benefit from screening.
	16. Do the results of this study fit with other available evidence?	No	Some prior evidence has been mixed though there is a lack of recent high-quality prospective comparative studies.
	17. What are the implications of this study for practice?	Unclear	Due to the various caveats above.
Overall judgement on methodological quality and applicability to question: overall low quality as a comparative study. It benefits from being an international registry with a large sample but otherwise there are extensive uncertainties and incomplete data.			

Appendix 4 – UK NSC reporting checklist for evidence summaries

All items on the UK NSC Reporting Checklist for Evidence Summaries have been addressed in this report. A summary of the checklist, along with the page or pages where each item can be found in this report, is presented in Table 28.

Table 28. UK NSC reporting checklist for evidence summaries

	Section	Item	Starting page no.
1.	TITLE AND SUMMARIES		
1.1	Title sheet	Identify the review as a UK NSC evidence summary.	Title page
1.2	Plain English summary	Plain English description of the executive summary.	5
1.3	Executive summary	Structured overview of the whole report. To include: the purpose/aim of the review; background; previous recommendations; findings and gaps in the evidence; recommendations on the screening that can or cannot be made on the basis of the review.	6
2.	INTRODUCTION AND APPROACH		
2.1	Background and objectives	Background – Current policy context and rationale for the current review – for example, reference to details of previous reviews, basis for current recommendation, recommendations made, gaps identified, drivers for new reviews	12
		Objectives – What are the questions the current evidence summary intends to answer? – statement of the key questions for the current evidence summary,	18

		criteria they address, and number of studies included per question, description of the overall results of the literature search. Method – briefly outline the rapid review methods used.	20
2.2	Eligibility for inclusion in the review	State all criteria for inclusion and exclusion of studies to the review clearly (PICO, dates, language, study type, publication type, publication status etc.) To be decided *a priori*.	22
2.3	Appraisal for quality/risk of bias tool	Details of tool/checklist used to assess quality, e.g. QUADAS 2, CASP, SIGN, AMSTAR.	24
3.	SEARCH STRATEGY AND STUDY SELECTION (FOR EACH KEY QUESTION)		
3.1	Databases/ sources searched	Give details of all databases searched (including platform/interface and coverage dates) and date of final search.	20
3.2	Search strategy and results	Present the full search strategy for at least one database (usually a version of Medline), including limits and search filters if used. Provide details of the total number of (results from each database searched), number of duplicates removed, and the final number of unique records to consider for inclusion.	51
3.3	Study selection	State the process for selecting studies – inclusion and exclusion criteria, number of studies screened by title/abstract and full text, number of reviewers, any cross checking carried out.	530
4.	STUDY LEVEL REPORTING OF RESULTS (FOR EACH KEY QUESTION)		
4.1	Study level reporting, results and risk of bias assessment	For each study, produce a table that includes the full citation and a summary of the data relevant to the question (for example, study size, PICO, follow-up period, outcomes reported, statistical analyses etc.).	Study level reporting: 59 Quality assessment: 77

		Provide a simple summary of key measures, effect estimates and confidence intervals for each study where available. For each study, present the results of any assessment of quality/risk of bias.	
5.	QUESTION LEVEL SYNTHESIS		
5.1	Description of the evidence	For each question, give numbers of studies screened, assessed for eligibility, and included in the review, with summary reasons for exclusion.	26, 33, 40
5.2	Combining and presenting the findings	Provide a balanced discussion of the body of evidence which avoids over reliance on one study or set of studies. Consideration of four components should inform the reviewer's judgement on whether the criterion is 'met', 'not met' or 'uncertain': quantity; quality; applicability and consistency.	29, 36, 44
5.3	Summary of findings	Provide a description of the evidence reviewed and included for each question, with reference to their eligibility for inclusion. Summarise the main findings including the quality/risk of bias issues for each question. Have the criteria addressed been 'met', 'not met' or 'uncertain'?	31, 38, 47
6.	REVIEW SUMMARY		
6.1	Conclusions and implications for policy	Do findings indicate whether screening should be recommended? Is further work warranted? Are there gaps in the evidence highlighted by the review?	48
6.2	Limitations	Discuss limitations of the available evidence and of the review methodology if relevant.	50

References

1. The UK NSC recommendation on Galactosaemia screening in newborns [Internet]. Public Health England. Available from: https://legacyscreening.phe.org.uk/galactosemia.
2. Pasquali M, Yu C, Coffee B. Laboratory diagnosis of galactosemia: a technical standard and guideline of the American College of Medical Genetics and Genomics (ACMG). Genet Med. 2018;20(1):3-11.
3. Seymour C, Thomason M, Chalmers R, et al. Newborn screening for inborn errors of metabolism: a systematic review. Database of Abstracts of Reviews of Effects (DARE): Quality-assessed Reviews [Internet]: Centre for Reviews and Dissemination (UK); 1997.
4. Varela-Lema L, Paz-Valinas L, Atienza-Merino G, et al. Appropriateness of newborn screening for classic galactosaemia: a systematic review. Journal of Inherited Metabolic Disease. 2016;39(5):633-49.
5. Welling L, Bernstein LE, Berry GT, et al. International clinical guideline for the management of classical galactosemia: diagnosis, treatment, and follow-up. Journal of Inherited Metabolic Disease. 2017;40(2):171-6.
6. Villoria JG, Pajares S, López RM, et al. Neonatal Screening for Inherited Metabolic Diseases in 2016. Seminars in Pediatric Neurology. 2016;23(4):257-72.
7. Hegarty R, Hadzic N, Gissen P, et al. Inherited metabolic disorders presenting as acute liver failure in newborns and young children: King's College Hospital experience. European Journal of Pediatrics. 2015;174(10):1387-92.
8. Ramadza DP, Sarnavka V, Vuković J, et al. Molecular basis and clinical presentation of classic galactosemia in a Croatian population. Journal of Pediatric Endocrinology and Metabolism. 2018;31(1):71-5.
9. Teke Kisa P, Kose M, Unal O, et al. Clinical and molecular characteristics and time of diagnosis of patients with classical galactosemia in an unscreened population in Turkey. Journal of Pediatric Endocrinology and Metabolism. 2019;32(7):675-81.
10. Porta F, Pagliardini S, Pagliardini V, et al. Newborn screening for galactosemia: a 30-year single center experience. World Journal of Pediatrics. 2015.
11. Viggiano E, Marabotti A, Burlina AP, et al. Clinical and molecular spectra in galactosemic patients from neonatal screening in northeastern Italy: Structural and functional characterization of new variations in the galactose-1-phosphate uridyltransferase (GALT) gene. Gene. 2015;559(2):112-8.
12. Welling L, Boelen A, Derks TGJ, et al. Nine years of newborn screening for classical galactosemia in the Netherlands: Effectiveness of screening methods, and identification of patients with previously unreported phenotypes. Molecular Genetics and Metabolism. 2017;120(3):223-8.
13. Lak R, Yazdizadeh B, Davari M, et al. Newborn screening for galactosaemia. Cochrane Database of Systematic Reviews. 2017;2017(12).
14. Özgün N, Celik M, Akdeniz O, et al. Early neurological complications in children with classical galactosemia and p.gln188arg mutation. International Journal of Developmental Neuroscience. 2019;78:92-7.
15. Rubio-Gozalbo ME, Haskovic M, Bosch AM, et al. The natural history of classic galactosemia: Lessons from the GalNet registry. Orphanet Journal of Rare Diseases. 2019;14(1).
16. Calderon FR PA, Crockett DK, Miller M, Mao R. Mutation database for the galactose-1-phosphate uridyltransferase (GALT) gene. Human Mutation. 2007;28(10):939-43.

17. Viggiano E, Marabotti A, Politano L, et al. Galactose-1-phosphate uridyltransferase deficiency: A literature review of the putative mechanisms of short and long-term complications and allelic variants. Clin Genet. 2018;93(2):206-15.
18. Van Calcar SC, Bernstein LE, Rohr FJ, et al. A re-evaluation of life-long severe galactose restriction for the nutrition management of classic galactosemia. Molecular Genetics and Metabolism. 2014;112(3):191-7.
19. Van Calcar SC, Bernstein LE, Rohr FJ, et al. Galactose content of legumes, caseinates, and some hard cheeses: implications for diet treatment of classic galactosemia. J Agric Food Chem. 2014;62(6):1397-402.
20. Honeyman M, Green A, Holton J, et al. Galactosaemia: results of the British Paediatric Surveillance Unit Study, 1988-90. Archives of disease in childhood. 1993;69(3):339-41.
21. Coelho AI, Rubio-Gozalbo ME, Vicente JB, et al. Sweet and sour: an update on classic galactosemia. J Inherit Metab Dis. 2017;40(3):325-42.
22. Murad MH, Sultan S, Haffar S, et al. Methodological quality and synthesis of case series and case reports. BMJ evidence-based medicine. 2018;23(2):60-3.