UK National Screening Committee

Newborn Screening for Mucopolysaccharidosis Type I

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

Version: FINAL

Author: Costello Medical

Date: October 2019

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 April 2020

Contents

Plain English summary

This review looked at screening in newborn babies for a rare genetic disorder called MPS I (mucopolysaccharidosis type I).

People with MPS I have a faulty version of an enzyme called alpha-L-iduronidase. This enzyme breaks down certain sugars in the body. When the enzyme does not work as it should (like in people with MPS I), these sugars can build up. This causes problems with children's physical and mental development.

Hurler syndrome is the more severe form of MPS I. Hurler-Scheie and Scheie syndromes are the less severe forms. Symptoms of Hurler syndrome often appear at around 6 months of age. Symptoms of Hurler-Scheie and Scheie syndromes develop later in childhood, typically between 3 and 10 years of age. MPS I worsens over time. It can lead to organ damage and early death.

Newborn screening might find babies with MPS I before symptoms appear. The UK National Screening Committee (UK NSC) last looked at the evidence for newborn screening for MPS I in 2014. The review found that there was not enough evidence to recommend a screening programme.

This current evidence summary updates the previous UK NSC review. It looks at all new evidence published since 2014. The focus of this review is to see:

- if current tests can accurately find babies with MPS I
- if early treatment is better for children with MPS I than later treatment, after symptoms appear

This review does not recommend screening for MPS I in newborn babies. This is because there is still not enough evidence:

- that screening newborn blood samples can accurately find MPS I
- that early treatment following screening will provide benefit to the baby or child

Executive summary

Purpose of the review

This review aimed to assess whether there have been significant developments in the evidence base since the last NSC review of newborn screening for mucopolysaccharidosis type I (MPS I). The purpose of this evidence synthesis was to assess whether the current UK NSC recommendation, not to implement screening for MPS I, should be reconsidered.

Background

MPS I is a rare, genetic lysosomal storage disorder caused by an autosomal recessive mutation in the α-L-iduronidase (*IDUA)* gene, leading to a deficiency of the IDUA enzyme responsible for degradation of the glycosaminoglycans (GAGs) heparan sulphate and dermatan sulphate. Traditionally, MPS I has been classified into 3 clinical phenotypes: Hurler syndrome (most severe), Hurler-Scheie syndrome, and Scheie syndrome (least severe).

Hurler syndrome patients appear normal at birth, and initially present with non-specific symptoms such as umbilical or inguinal hernia that are then typically followed by progressive symptoms including skeletal dysplasia, intellectual disability, hepatic disease, cardiorespiratory and central nervous system deterioration, and hearing loss. If left untreated, Hurler syndrome patients are unlikely to survive beyond 10 years of age. Symptoms are usually less severe for Hurler-Scheie and Scheie patients; although mild or no cognitive impairment is experienced, life expectancy can still be substantially reduced. Advancements to available treatment options such as haematopoietic stem cell transplantation (HSCT) and enzyme replacement therapy (ERT) allow patients to experience improvements in these clinical symptoms.

Globally, the median age at diagnosis for MPS I patients is reported to be between 1 and 5 years of age, with UK patients diagnosed around 5 years of age. While early diagnosis is considered to be important to allow for early treatment, particularly for patients with Hurler syndrome, this can be difficult as the first clinical symptoms are unspecific. Published guidelines are relatively consistent on a proposed diagnostic algorithm that captures the stages of analytical testing necessary to confirm an MPS I diagnosis in newborns who have screened positive for MPS I. The stages typically include measurement of leukocyte IDUA enzyme activity followed by gene *IDUA* molecular analysis and urinary GAG testing. The suggested algorithm also reflects the recommended pathways for MPS I treatment or monitoring, which depend on diagnostic status.

Nevertheless, the last UK NSC review reported that the impact of early initiation of treatment is yet to be determined, and it is unclear whether a newborn MPS I screening programme would result in clinical benefits for MPS I patients.

Focus of the review

This review aimed to evaluate whether the evidence base has developed substantially and a screening programme for MPS I has become viable since the previous UK NSC review was conducted in 2014. Specifically, new evidence was collected to answer the following 2 questions:

- What is the accuracy of commercially available screening tests in dried blood spots (DBS) to detect MPS I? (criterion 4)
- Does early initiation of treatment with HSCT and/or ERT following screening provide better outcomes compared to usual clinical care? (criterion 9)

Recommendation under review

The UK NSC has previously considered evidence for screening for MPS I in 2015. Based on evidence identified by the literature review conducted in 2014, screening for MPS I in newborns in the UK was not recommended. The review concluded that the evidence base was limited in volume, quality and consistency regarding the performance of available testing strategies for newborn screening, and any potential benefit of early treatment following screen detection (or an optimum age for treatment initiation).

Findings and gaps in the evidence of this review

Seventeen publications were extracted and included in the evidence synthesis. A summary of question level results is presented below.

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

Only 4 studies were published since the previous review (2014). Based on the evidence assessed by this review, there is limited evidence relating to the accuracy of newborn DBS screening tests for MPS I. Three studies measured IDUA enzymatic activity by tandem mass spectrometry, and one study evaluated a fluorometric assay in combination with a pattern recognition software. However, there was substantial heterogeneity in screening test methods and lack of reported measures of test performance, with only positive predictive values (PPVs) reported. Based on these findings, criterion 4 is 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. Evidence relating to wider benefits of screening, for example those relating to family members, should be taken into account where available. However, where there is no prospect of benefit for the individual screened then the screening programme should not be further considered.'

Thirteen studies evaluated the relationship between age at initiation of HSCT or ERT and clinical outcomes for MPS I patients. Although some indicated a statistically significant association, the effect was small, and it is therefore unclear whether early diagnosis of MPS I would result in a clinically significant improvement in patients' symptoms. Other studies did not demonstrate any effect of age of treatment initiation on clinical outcomes. The majority of studies focussed solely on Hurler patients, while the effect of early initiation of treatment for patients with attenuated MPS I was rarely investigated. The median age of treatment in these studies was also more aligned with clinical detection of MPS I rather than earlier initiation of treatment following detection through screening. Overall, there is insufficient evidence to determine whether early initiation of HSCT or ERT improves clinical outcomes for MPS I patients. Based on these findings, criterion 9 is not met.

Evidence uncertainties

Criterion 4 – Full reporting of test accuracy for assessment of IDUA activity by tandem mass spectrometry or fluorometric assay was lacking in all studies; only PPV was reported without any measure of variance, diminishing confidence in the results. In all newborn screening programmes identified, only screen-positive samples were sent for confirmatory testing. Though further evidence on the sensitivity and specificity of tests to detect MPS I would be ideal, it is acknowledged that assessment of these test accuracy parameters is difficult to achieve in studies of screening for rare diseases. Finding mechanisms to address that is important, particularly given the potential for identification of carriers and pseudodeficiency in MPS I screening. Further screening studies with improved methodological consistency (in terms of index test cut-offs, repeat testing and the reference standard used) may be achievable and would allow for an informative evaluation of a putative test to be used in screening for MPS I in newborn babies.

Criterion 9 – Although there may be clinical benefits in initiating HSCT early in patients diagnosed with Hurler syndrome, the evidence was inconsistent in terms of the treatment outcomes which were investigated, results across similar outcomes, and the age by which early and late treatment groups were defined. In some cases, age was only analysed as a

covariate of the outcome measured, further limiting the conclusions that can be drawn. The evidence is also limited by differences in study design, methodology, and small sample sizes across the included studies. Further consistency in study methodology and investigated treatment outcomes in future research may mitigate some of the current issues present within the current evidence base that are inherent in rare diseases such as small sample sizes.

Recommendations on screening

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

There was insufficient evidence to determine whether newborn DBS screening using tandem mass spectrometry or fluorometric assays is sufficiently accurate to identify all patients with MPS I. While there is some evidence from studies at high risk of bias that early HSCT may improve treatment outcomes in patients with Hurler syndrome specifically, ultimately no clear conclusions can be drawn on whether this provides any benefit for affected cases and their families.

Limitations

This review only included peer-reviewed journal publications in the English language. Given that this is an accepted methodological adjustment for a rapid review and this review was focusing on evidence relevant to the UK setting, these limitations should not have led to the exclusion of any pivotal studies. The titles, abstracts and full texts were screened by one reviewer, with a second reviewer verifying all included, 10% of excluded decisions and any articles where there was uncertainty about their inclusion. For question 1, publications were excluded if they only presented data that would require a calculation of test accuracy parameters that were otherwise not reported. This was taken as a pragmatic approach and was unlikely to result in key screening studies being missed.

Introduction and approach

Background

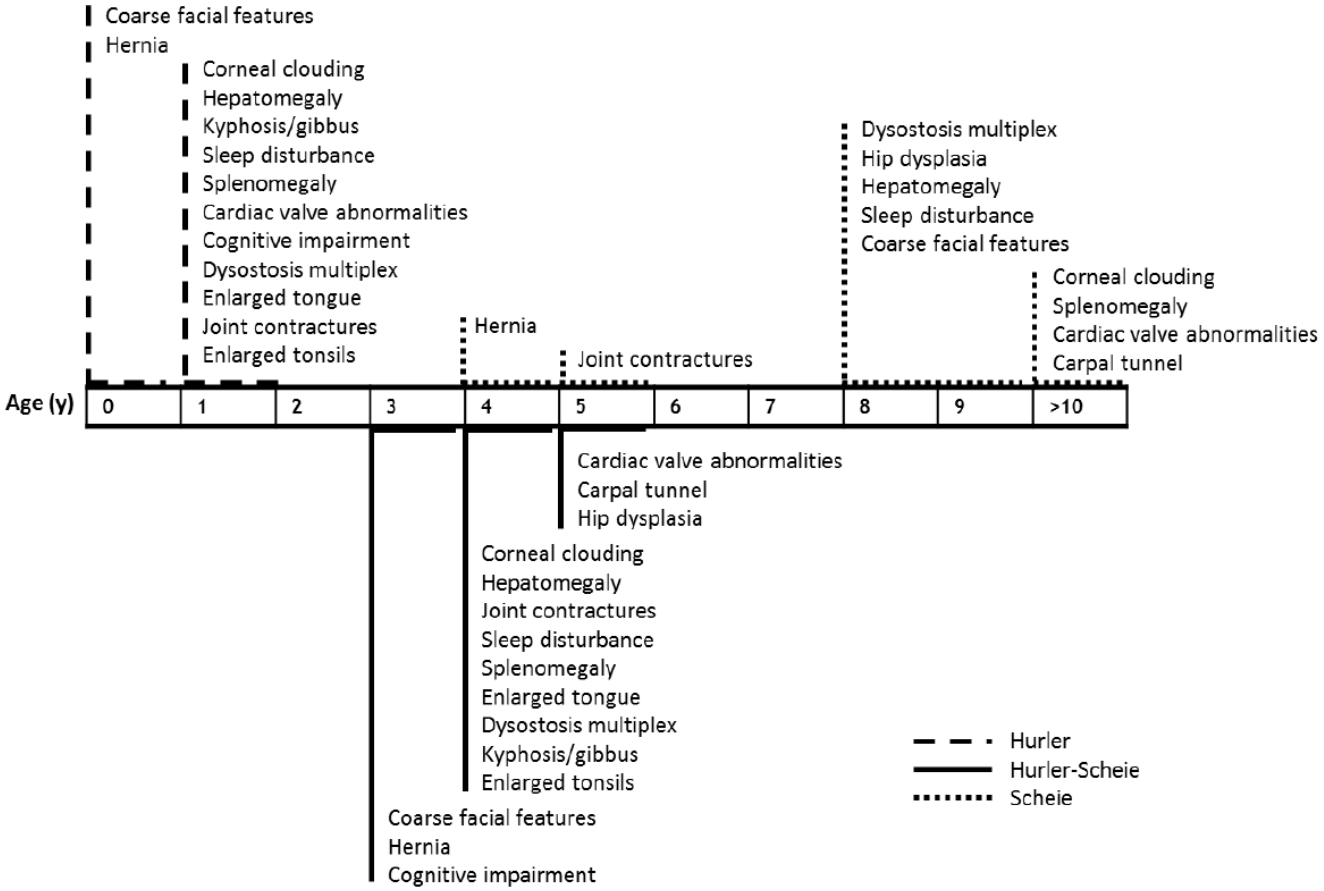
Mucopolysaccharidosis type I (MPS I) is a rare, genetic lysosomal storage disorder (LSD) caused by an autosomal recessive mutation in the α-L-iduronidase (*IDUA)* gene, leading to a deficiency of the IDUA enzyme responsible for degradation of glycosaminoglycans (GAGs), heparan sulphate (HS) and dermatan sulphate (DS). Subsequent accumulation of GAGs results in progressive multi-organ deterioration, symptoms of which can vary widely in terms of the timing of presentation and severity, with the most severe cases resulting in early death.[1-3]

Traditionally, MPS I has been classified into 3 clinical phenotypes: Hurler syndrome, Scheie syndrome, and Hurler-Scheie syndrome. Hurler syndrome is the most severe form of MPS I, causing symptoms to appear early in life and progress rapidly in severity. By contrast, Scheie syndrome manifests later and displays slower disease progression, with Hurler-Scheie patients being mostly in between the other 2 in terms of disease onset and severity. Nevertheless, due to biochemical and closer clinical overlap in phenotype, the Scheie and Hurler-Scheie subtypes are now more commonly referred to as 'attenuated MPS I'.[4]

Infants with Hurler syndrome will appear normal at birth and initially present with non-specific symptoms, such as umbilical or inguinal hernia and upper respiratory-tract infections, at a median age of 6 months (Figure 1).[3, 4] Gibbus deformity and coarsening of facial features may also develop. These are typically followed by progressive skeletal dysplasia, intellectual disability, hepatic disease, cardiorespiratory and central nervous system deterioration and hearing loss, with linear growth also decreasing.[3, 4] If left untreated, Hurler syndrome patients are unlikely to survive beyond 10 years of age.[3, 4]

For patients with attenuated MPS I, the onset of symptoms is generally after infancy, between 3 and 10 years of age (Figure 1).[3, 4] Patients with Scheie syndrome typically have normal cognitive functioning and survive into adulthood, although more than 50% experience cardiac valve abnormalities, hernias, corneal clouding, and hepatomegaly.[3] Hurler-Scheie patients present with an intermediate phenotype, showing mild or no cognitive impairment, but exhibiting symptoms that ultimately reduce life expectancy to less than 40 years of age.[3]

Figure 1. Timeline for symptom onset among Hurler, Hurler-Scheie and Scheie patients based on MPS I registry data



Reproduced from the previous UK NSC (2015) report on newborn screening for mucopolysaccharidosis I.[5] This timeline was based on MPS registry data.[3, 6-8]

Advancements to available treatment options such as haematopoietic stem cell transplantation (HSCT; also known as bone and marrow transplantation [BMT]) and enzyme replacement therapy (ERT), allow patients to experience improvements in these clinical symptoms.[3]

The UK National Screening Committee (UK NSC) wishes to determine whether there have been significant developments in the evidence base on how earlier diagnosis and subsequent initiation of treatment may be beneficial for MPS I patients, and if a UK national screening programme may now be a viable option. In addition to this, the UK NSC is interested in contextualising this review by investigating the usual age of clinical diagnosis for MPS I and whether a diagnostic pathway has been established for pre-symptomatic or screen detected MPS I patients.

Usual age of clinical diagnosis for MPS I

The age at which MPS I is diagnosed can vary substantially, both across and within the 3 generally recognised forms of MPS I: Hurler, Scheie, and Hurler-Scheie syndromes. While early diagnosis is considered to be important to allow for early treatment, particularly for patients with Hurler syndrome, this can be difficult as the first clinical symptoms are unspecific.[9, 10] It is worthwhile noting that there is significant heterogeneity in the age when MPS I is first diagnosed among individual patients, with some reports differing by up to 54.1 years,[6] likely indicating that for many patients, establishing a diagnosis is difficult. Globally, reported median age at diagnosis for MPS I (when not further differentiated into subtypes) varies between 1 and 5 years of age, with UK patients diagnosed closer to the top of this range (5 years of age).[11] However, when different clinical phenotypes are considered, the age at diagnosis appears to vary significantly; for patients with the more severe Hurler phenotype it is substantially lower, with median age at diagnosis across the world ranging from 0.8 to 1.0 year of age, but is between 7.0 and 9.4 years for those with the attenuated Scheie phenotype. Unsurprisingly, the median age at diagnosis in Hurler-Scheie patients falls in between these values, ranging from 3.8 to 4.0 years of age.

Data on how geographical areas compare as to the age at diagnosis is scarce; based on what is available, the pattern that attenuated MPS I is diagnosed later than Hurler syndrome is similar across different regions of the world.[3] There is, however, variation between geographical areas, with the age of diagnosis for Hurler and Hurler-Scheie syndromes being lower in Europe than in other regions, and, conversely, Scheie patients having the lowest age of diagnosis in North America.[3] For the UK specifically, there is limited data on age at diagnosis; available reports suggest that the median age at diagnosis for Scheie patients (7 years of age) is lower in the UK than other European countries, America or Asia, whilst for Hurler-Scheie patients the median age at diagnosis in the UK (4 years of age) is slightly higher than other European countries, though lower than in other areas of the world.[11] Mean age of Hurler syndrome diagnosis appears to be around 9 months in the UK, which is comparable to the global averages of 0.8 to 1.0 years.[12]

Diagnostic pathway for pre-symptomatic (e.g. cascade testing of siblings) or screen detected MPS I

As MPS I is a genetic condition, siblings of a diagnosed child could also be affected.[13] Case study reports have demonstrated that MPS I can be diagnosed early after birth when the test for the disorder was prompted by the diagnosis of a sibling. For example, MPS I cases detected via a sibling cascade have been reported to have been diagnosed and classified as attenuated and Hurler syndrome forms as early as 3 and 10 days old, respectively. This

allowed for initiation of treatment before symptoms arose in both cases (at 5 months of age for attenuated MPS I and at 2 months of age for Hurler syndrome).[14, 15]

Although sibling-cascade or carrier testing has been shown to be useful for detecting cases of MPS I at an earlier age,[13] there are currently no agreed national guidelines for cascade testing of siblings. Nevertheless, when a newly diagnosed MPS I patient is evaluated for treatment, *IDUA* sequencing of at-risk family members has been suggested.[16]

MPS I can be diagnosed prenatally by chronic villus sampling or amniocentesis, using relevant biochemical and molecular tests, however no formal recommendations have been made regarding prenatal testing.[17] By contrast, several guidelines and recommendations have been published on the diagnostic pathway for patients who have screened positive for MPS I in the newborn period.[16, 18] These guidelines are relatively consistent on a proposed diagnostic algorithm that captures the stages of analytical testing necessary to confirm an MPS I diagnosis (adapted in Figure 2). The suggested algorithm also reflects the recommended pathways for MPS I treatment or monitoring, which depend on diagnostic status.

Even in individuals known to have MPS I, prediction of clinical phenotype can be difficult. It is generally agreed that infants with a positive MPS I screening test result should undergo follow-up testing for plasma, peripheral blood leukocyte or skin fibroblast IDUA enzyme activity, to confirm the test result and classify the diagnosis as the expected clinical severity. No residual enzyme activity will typically be found in severe MPS I, whilst patients with the attenuated Scheie phenotype will usually present with <0.1% of the normal level of activity.[11] However, phenotypes cannot be reliably differentiated based on IDUA enzyme assay only, as reduced *in vitro* IDUA activity can also be observed with benign variants that do not cause symptoms in the individual (referred to as pseudodeficiency).[18] Kingma 2013, a retrospective analysis, reported on a diagnostic algorithm involving enzymatic analysis of IDUA activity in fibroblasts that allowed for differentiation between Hurler syndrome and attenuated MPS I in a sample of 30 affected newborns with 82% sensitivity and 100% specificity, the performance of this algorithm is yet to be validated in a prospective cohort.[2] As there is not a validated and reliable method to classify MPS I diagnoses, and enzymatic activity alone is insufficient, infants with low IDUA activity (<1% of the normal)[19] should be referred to a metabolic centre for molecular analysis. Correlations between recurrent pathogenic mutations and phenotype of MPS I have been identified, which could, in some cases, allow for prediction of phenotype before symptoms arise. For example, pre-symptomatic newborns with IDUA deficiency and known severe *IDUA* mutations (e.g. nonsense common W402X and Q70X, missense A327P and G51D) on both alleles, have been reported to develop severe MPS I due to lack of a functional enzyme.[20] *IDUA* variant analyses indicate that pre-symptomatic newborns with attenuated disease have at least 1

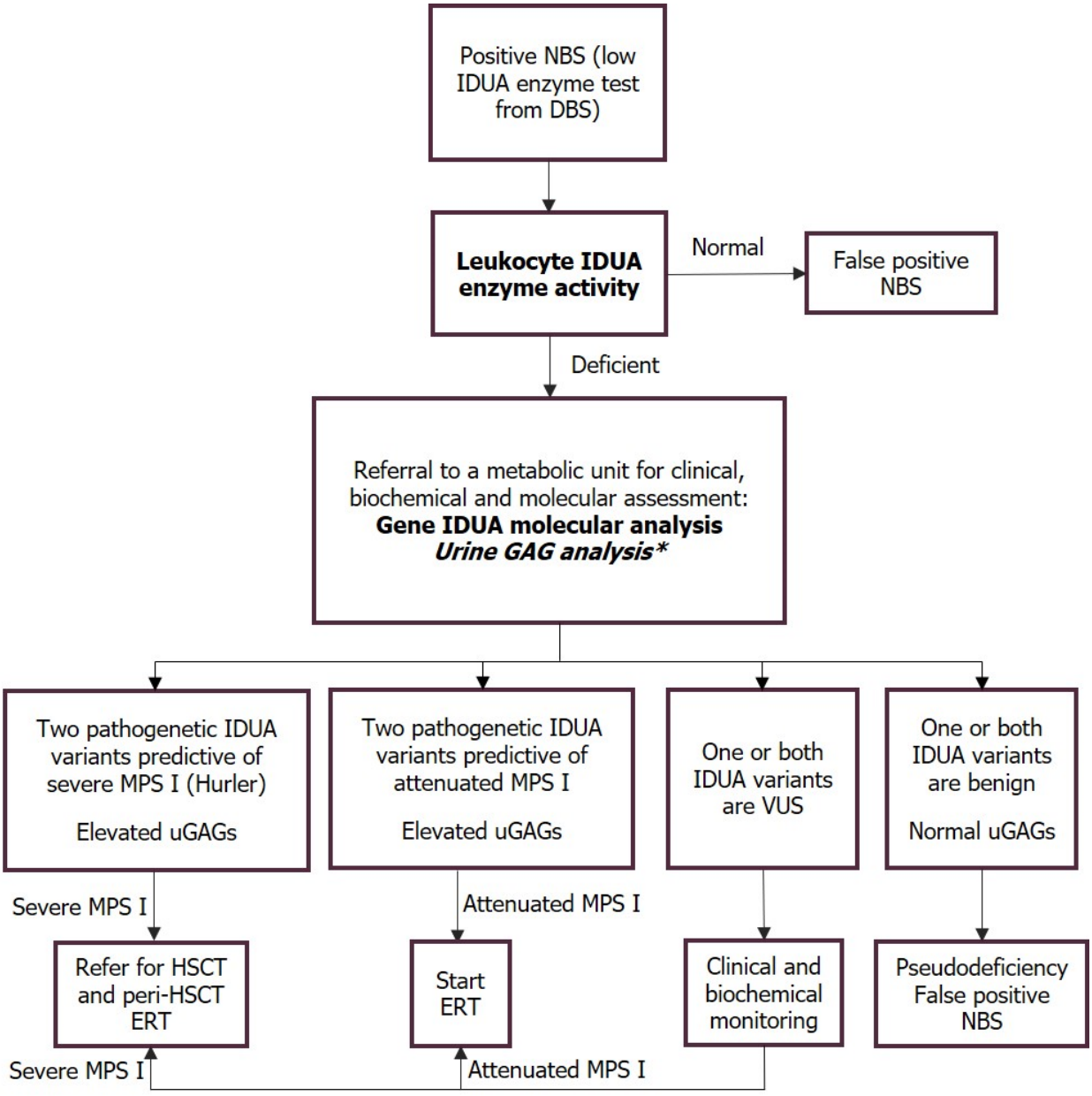
allele containing a missense or splice site variant (e.g. R89W and L492P). However, the phenotype of MPS I cannot be predicted for all genotypes, such as for less common or unique mutations in individuals and families.[18]

When the pathogenic variants can be used to predict disease phenotype, the treatment pathways are clearly described, and the recommendation is that these should be initiated promptly; if 2 pathogenic alleles associated with severe disease have been identified and urinary GAG (uGAG) levels are elevated, current recommendations suggest that the patient should be referred for HSCT as soon as possible (Figure 2). If the pathogenic variants identified indicate attenuated disease, then ERT should be initiated (Figure 2).[18, 20] Nonetheless, there are circumstances where disease severity cannot be predicted, due to the presence of variants of unknown significance (VUS). In these cases, ongoing monitoring through physical examinations, clinical evaluations and biochemical analyses may help to anticipate disease severity in order to determine appropriate treatment pathways (Figure 2). It is recommended that patients should be assessed for facial dysmorphisms, joint range of motion, murmurs, liver and spleen enlargement, corneal clouding, hernias, and scoliosis/kyphosis; while these symptoms may not arise until later in childhood,[2] awareness and monitoring for clinical presentation may facilitate the diagnostic process. Clinical evaluations including radiographs, echocardiography, neurocognitive testing, ophthalmologic assessment and magnetic resonance imaging (MRI), are also recommended. Findings of gibbus deformity, dysostosis multiplex, corneal/cardiac valve involvement, and respiratory distress indicate a severe disease phenotype. If the phenotype of MPS I cannot be predicted, these assessments are recommended at 3 and 6 months of age, and if inconclusive they should be continued on an ongoing basis.[18] These ongoing evaluations aim to help predict disease phenotype and determine the appropriate treatment regimen according to the diagnostic algorithm (Figure 2). However, if conclusive classification of disease subtype cannot be achieved by at least 6 months of age, it is suggested that ERT should be started regardless. Nevertheless, it is important to consider that any symptoms reflective of severe MPS I may be masked if ERT is started early, which may result in delayed initiation of HSCT for these patients, demonstrating the complexity of therapeutic decision-making in infants with MPS I.[18] These issues highlight the importance of obtaining a method that can reliably distinguish MPS I patients who will gain optimal benefit from initiating HSCT at an early age from those who will benefit from receiving ERT.[2]

Although informative, uGAG tests revealing excessive dermatan and heparan sulphate cannot be used to confirm MPS I diagnosis, as elevation of uGAGs is not specific to MPS I.[21] Furthermore, although patients with Hurler syndrome typically have higher uGAG levels than those with attenuated disease, the exact threshold that distinguishes these subtypes has not been determined; particularly in newborns for whom the normative uGAG ranges

are relatively large.[18] The importance of uGAG analysis is therefore not well agreed upon, and whilst some suggest it should be the first step in the diagnostic pathway,[11] others recommend performing it once IDUA activity has been confirmed (Figure 2).[16, 18, 22] It is recommended that quantitative and qualitative GAG analyses should be interpreted together for a complete evaluation.[23]

Figure 2. Proposed diagnostic algorithm for positive MPS I newborn screening



Adapted from Clarke et al. (2016), Donati et al. (2018) and Wang et al. (2011).[16, 18, 22] *There is no consensus as to whether uGAG analysis should be performed as the first step in the diagnostic pathway after a positive NBS, or alongside molecular analysis following leukocyte IDUA enzyme activity analysis. DBS: dried blood spot; ERT: enzyme replacement therapy; HSCT: haematopoietic stem cell transplantation; IDUA: alpha-L-iduronidase; MPS I: mucopolysaccharidosis type I; NBS: newborn screening; uGAG: urine glycosaminoglycan; VUS: variant of unknown significance.

Current global landscape of newborn screening

Several countries have already initiated newborn screening (NBS) programmes for MPS I to facilitate early diagnosis of affected newborns. In the US, the Department of Health and Human Services Secretary's Advisory Committee on Heritable Disorders in Newborns and Children recommend that MPS I is included in the newborn screening panel.[24] However, whilst each state must offer screening for every infant, the choice to include MPS I depends on the individual state's public health department; pilot studies have so far been initiated in, for example, Illinois, Washington, and Missouri.[25-27] Other countries including Taiwan, Brazil and in the Tuscany and Umbria regions of Italy, are also conducting pilot studies for MPS I screening programmes.[28-30] Furthermore, in the North-East region of Italy and in Mexico, MPS I has been included in the NBS programmes for detecting several LSDs in neonates.[31, 32]

Current policy context and previous reviews

The UK NSC has considered evidence in support of screening for MPS I in 2015.[5] A literature review was conducted in 2014 and it concluded that the evidence base was limited in volume, quality, and consistency. As such, screening for MPS I in newborns in the UK is currently not recommended.

Objectives

This review aims to assess whether there have been significant developments in the evidence base since the previous review was conducted in 2014, and if sufficient support exists for a screening programme for MPS I in newborn infants. The review appraised evidence on the questions in Table 1, which each relate to 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		
	THE TEST		
4	There should be a simple, safe, precise and validated screening test.	What is the accuracy of commercially available screening tests in dried blood spots (DBS) to detect MPS I?	4 publications on 4 unique cohorts
	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.	Does early initiation of treatment with haematopoietic stem cell transplantation (HSCT) and/or enzyme replacement therapy (ERT) following screening (i.e. universal newborn screening and cascade testing of siblings) provide better outcomes compared to usual clinical care?	13 publications on 13 unique cohorts[a]

[a] Two publications included the same medical centres as sources of patients, however, it was not possible to determine whether the patients overlapped. The studies were treated as being on independent cohorts.

Methods

The current review was conducted by Costello Medical, in keeping with the UK National Screening Committee evidence review process. Database searches were conducted on 2 May 2019 to identify studies relevant to the questions detailed in Table 1.

Eligibility for inclusion in the review

The following review process was followed:

- Each abstract was reviewed against the inclusion/exclusion criteria by one reviewer (Sift 1). Where the applicability of the inclusion criteria was unclear, the article was included at this stage to ensure that all potentially relevant studies were captured. A second independent reviewer provided input in cases of uncertainty and validated all included and 10% of excluded articles. Any disagreements were resolved by discussion until a consensus was met.
- Costello Medical conducted a search for freely available full-text articles required for the full-text review stage (Sift 2) and acquired any additional articles from the Cambridge University Library.
- Each full-text article was then reviewed against the inclusion/exclusion criteria by one reviewer (Sift 2), who determined whether the article was relevant to one or more of the review questions. A second independent reviewer provided input in cases of uncertainty and validated all included and 10% of excluded articles. Any disagreements were resolved by discussion until a consensus was met.

Eligibility criteria for each question are presented in Table 2 (Question 1) and Table 3 (Question 2) below:

Table 2. Inclusion and exclusion criteria for question 1 — What is the accuracy of commercially available screening tests in dried blood spots (DBS) to detect MPS I?

Domain	Population	Target condition	Intervention	Reference Standard	Outcome	Study type	Study setting	Other considerations
Inclusion criteria	Newborn infants	MPS Type I	**Index test**: • IDUA activity tested by tandem mass spectrometry (MS/MS) • IDUA activity tested via fluorometric enzyme assay • Assays aiming to detect lysosomal storage disorders that include MPS I • Other tests considered on case by case basis	• IDUA activity measured in leukocytes from a whole blood sample • Molecular DNA analysis • Urine glycosaminoglycan (GAG) quantification • Other reference standards to be considered on case by case basis	Measures of screening accuracy: • Sensitivity • Specificity • Positive predictive value • Negative predictive value • Accuracy • Likelihood ratio	**Tier 1:** RCTs and interventional studies, cross-sectional studies, cohort studies, systematic reviews and meta-analyses **Tier 2**: Case-control studies *Tier 2 evidence was eligible due to insufficient studies at Tier 1*	**Tier 1**: Studies conducted in the UK **Tier 2**: Studies conducted in EEA or OECD countries **Tier 3**: Studies conducted in other countries *Tier 2 and 3 evidence was eligible for inclusion due to a limited number of studies conducted in the UK*	Peer-reviewed studies in the English language Studies published since 2014
Exclusion criteria	• Infants over 28 days of age or		Studies not using a combination of an index test and a reference standard		Any other outcomes (including area under the receiver-	Case reports, case series, narrative reviews, editorials,		Studies with full text not in the English language

	older children, adults, pregnant women • Mixed populations where results are not presented separately for infants				operator curve or measures of association between test outcome and MPS I)	commentaries, letters, conference abstracts or other publication types that have not been peer-reviewed		Studies published pre-2014

Abbreviations: GAG, glycosaminoglycan; IDUA, Iduronidase; MPS I, Mucopolysaccharidosis type I; RCT, randomised controlled trial

Table 3. Inclusion and exclusion criteria for question 2 — Does early initiation of treatment with HSCT and/or ERT following screening provide better outcomes compared to usual clinical care?

Domain	Population	Target condition	Intervention	Comparators	Outcome	Study type	Study setting	Other considerations
Inclusion criteria	Patients with MPS I	MPS Type I	• Early HSCT • ERT Used alone or in combination	**Tier 1:** Late HSCT and/or ERT **Tier 2:** No treatment or placebo *Tier 2 evidence was eligible due to insufficient studies at Tier 1*	Efficacy and safety outcomes including but not limited to: • Mortality • Cognitive development • Growth • Orthopaedic outcomes • Cardiac	RCTs and interventional studies, cohort studies, systematic reviews and meta-analyses	**Tier 1**: Studies conducted in the UK **Tier 2**: Studies conducted in EEA or OECD countries	Peer-reviewed studies in the English language Studies published since 2014

				• outcomes • Respiratory outcomes • Ophthalmologic outcomes • Hearing impairment • Overtreatment (e.g. in cases with no clinical symptoms/pseudodeficiency) • Quality of life outcomes		**<u>Tier 3</u>**: Studies conducted in other countries *Tier 2 and 3 evidence was eligible for inclusion due to a limited number of studies conducted in the UK*	
Exclusion criteria	Patients without MPS I or populations where results for MPS I patients are not presented separately	Any other interventions	Any other comparator	Any economic outcomes e.g. costs, length of hospital stay	Case reports, case series, narrative reviews, editorials, commentaries, letters, conference abstracts or other publication types that have not been peer-reviewed		Studies with full text not in the English language Studies published pre-2014

Abbreviations: ERT, enzyme replacement therapy; HSCT, haematopoietic stem cell transplantation; MPS I, mucopolysaccharidosis Type I

Appraisal for quality/risk of bias tool

The following tools were used to assess the quality and risk of bias of each study included in the review:

- Diagnostic accuracy studies: Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool
- Risk of Bias in Non-Randomised Studies of Interventions (ROBINS-I)

Results of the quality assessments and appraisal of individual studies are presented in Appendix 3.

Databases/sources searched

The following databases were searched:

- MEDLINE, MEDLINE In-Process, MEDLINE Daily, Epub Ahead of Print
- Embase
- The Cochrane Library, including the following:
    - Cochrane Database of Systematic Reviews (CDSR)
    - Cochrane Central Register of Controlled Trials (CENTRAL)
- Database of Abstracts of Reviews of Effects (DARE)

Searches were conducted in May 2019. Full details of the searches, including the search strategy for each database, are presented in Appendix 1.

Question level synthesis

Criterion 4 — What is the accuracy of commercially available screening tests in dried blood spots to detect MPS I?

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

Question 1 – What is the accuracy of commercially available screening tests in dried blood spots (DBS) to detect MPS I?

The previous (2015) UK NSC review evaluated evidence relating to the clinical value of newborn screening tests for MPS I and identified a limited volume of evidence relating to 2 screening test strategies for assessing IDUA activity (fluorometric enzyme assay and tandem mass spectrometry). Studies were of poor quality and measures of test accuracy such as sensitivity, specificity, and likelihood ratios were not reported, limiting the evaluation of the performance and clinical validity of these tests for newborn screening. The 2015 review concluded that further research on the sensitivity, specificity, positive and negative predictive values of various MPS I testing strategies was required in order to assess the effectiveness of various screening tests for MPS I.

The aim of this question was to identify and synthesise evidence published since 2014 on test accuracy parameters of screening tests for MPS I in newborn infants in a low-risk or unselected UK population.

Eligibility for inclusion in the review

This review searched for randomised controlled trials (RCTs), cohort, cross-sectional, and interventional studies with an appropriate screening component, along with systematic literature reviews (SLRs) or meta-analyses (MAs). As limited evidence was identified, case-control studies and studies conducted in any country were eligible for inclusion. Studies were eligible if they assessed the performance of an index test used to diagnose MPS I in newborn infants, such as measurement of IDUA activity (by mass spectrometry or enzyme assay) or assays aimed at detecting LSDs including MPS I. The reference standard was IDUA activity measured in leukocytes (blood sample), molecular DNA analysis or urinary GAG quantification. Studies were only included if they directly reported test accuracy parameters; no calculations were performed in this review to obtain such measures. The eligible population were newborn infants younger than 28 days of age; studies that

evaluated screening for MPS I in infants over 28 days of age or older children, adults or pregnant women only were not included.

Full details of the eligibility criteria are presented in Table 2.

Description of the evidence

A total of 4 publications were included in the review for Criterion 4. No SLRs or MAs which aligned with the scope of this review closely enough to be included in their own right were identified, and no additional relevant articles were identified through hand-searching the reference lists of the identified SLRs. No studies investigating index tests or reference standards other than those specified in the eligibility criteria were identified.

The 4 included studies reported on newborn screening programmes in the US,[25, 35] Italy,[31] and Taiwan.[36] Three studies reported on combined screening for 6 LSDs including MPS I in 43,701 to 55,161 newborns.[25, 31, 35] The remaining study reported on a screening programme for MPS I and II, during which 294,196 and 153,032 newborns were tested for MPS I or MPS II respectively.[36] None of the studies reported screening for a specific MPS I phenotype.

Three studies assessed IDUA activity measured by tandem mass spectrometry,[31, 35, 36] and 1 study by fluorometric enzymatic assay,[25] in newborn DBS as an index screening test for MPS I. Newborn DBS samples were collected between 36 and 58 hours after birth. In 2 studies the time of DBS sample collection was unclear; in Chuang 2018 the age at which confirmatory genetic testing was performed was reported for the 4 newborns ultimately diagnosed with MPS (between 1.1 months and 4.5 months after birth).[25, 36] In all 4 studies, the screening phase involved 2 or more stages of testing. In the Burlina 2018 screening programme conducted in Italy, a repeat index test was performed using a second newborn DBS sample for those who initially screened positive (cut-off <0.2 MoM µmol/h). Only those newborns who tested positive again, using the same cut-off, received the reference standard for confirmation.[31] Similarly, Chuang 2018 re-tested the original DBS and a second DBS, using the same cut-off for each round of testing.[36] Newborns who consistently tested positive (cut-off <4.0 µmol/L/hour [0.8 percentile]) with triplicate testing were referred for confirmatory testing in the Hopkins 2015 screening study. For ‘high-risk’ newborns in whom screening results were considered unreliable (e.g. in premature or unwell newborns, or those less than 24 hours old), a repeat screen was automatically required.[25] By contrast, Minter Baerg 2018 used a ‘second-tier’ test in addition to a repeat screen of IDUA activity in DBS samples – evaluation of dermatan sulphate and heparan sulphate concentrations – before referring newborns for confirmatory testing.[35]

Two studies reported the use of pre-specified cut-offs for classification of a 'positive' or 'negative' index test result for MPS I, determined using samples from healthy patients and clinical cases of MPS I.[25, 31] To distinguish newborns at risk (positive test result) or not at risk (negative test result), Minter Baerg 2018 reported a thorough screening process which utilised pattern recognition software (The Collaborative Laboratory Integrated Reports [CLIR] tool). While numerical cut-offs were not explicitly pre-specified, it was reported that the initial index test was set to only classify samples as 'negative' if the 6-plex assay results were 'completely normal' (if markers were within 1–99% percentile of reference range).[35]

Where reported, more than one test was used as the reference standard for MPS I across the 4 newborn screening programmes. For example, a leukocyte assay for β-iduronidase to confirm IDUA activity, and molecular DNA analysis using polymerase chain reaction (PCR) for confirmation of genotype, were performed in the newborn screening programme in Taiwan.[36] In the Regional North East Italy screening programme, clinical evaluation, urinary GAG analyses, further IDUA testing and genetic mutational analyses were performed to confirm MPS I and classify the phenotype as Hurler, Scheie or Hurler-Scheie.[31] No other studies reported on the MPS I phenotype despite the use of genetic testing. The diagnostic laboratory tests used to confirm diagnoses were not explicitly reported by Hopkins 2015, however samples that tested positive using the index test were referred to genetic centres for 'evaluation, confirmatory testing and diagnosis'.[25]

Discussion of findings

A study-level summary of data extracted from each included publication is presented in Table 5.

Quality assessment

The quality of the included studies was appraised using an adapted QUADAS-2 checklist (Table 17). A summary of the risk of bias and applicability to the UK setting is presented in Table 4, and the full appraisal is presented in Table 18; Appendix 3.

Table 4. Summary of QUADAS-2 assessments for MPS I screening studies

Question	Burlina 2018[31]	Chaung 2018[36]	Hopkins 2015[25]	Minter Baerg 2018[35]
PARTICIPANT SELECTION				
Risk of bias	Low	Low	Low	Low
Concern about applicability	Low	High	Low	Low
INDEX TESTS				
Risk of bias	Low	Low	Low	Unclear
Concern about applicability	High	Unclear	Unclear	High
REFERENCE STANDARD				
Risk of bias	Low	Low	High	Low
Concern about applicability	Low	High	High	High
PARTICIPANT FLOW				
Risk of bias	High	High	High	High

Participant selection

All studies reported on unselective newborn screening programmes and were therefore at low risk of bias. Chuang 2018 reported on a newborn screening programme in Taiwan, limiting applicability to the UK newborn screening population and UK clinical practice.[36] The remaining 3 screening studies were conducted in the US and Italy, where the newborn screening population and clinical practice is considered likely to be comparable to the UK, resulting in low concern about applicability.[25, 31, 35]

Index tests

In all screening programmes, only newborns who screened positive using the index test were referred for confirmatory testing with the reference standard, which suggests that the index test results were interpreted without the knowledge of the reference standard. Three studies reported the use of pre-specified cut-offs for classification of a 'positive' or 'negative' test result for MPS I, and are therefore at low risk of bias. In the Minter Baerg 2018 study, while the initial index test using the CLIR tool was set to only give a negative test result if the 6-plex assay results were 'completely normal' (markers within 1–99% percentile of reference range),[35] the final cut-off used to identify screen-positive newborns for confirmatory testing was unclear. The CLIR tool was developed to improve screening by tandem mass spectrometry and provide continuous, covariate-adjusted, 'moving' percentiles, but as the study did not report which thresholds were used specifically, the study is at an unclear risk of bias.

The North Eastern Italy screening programme collected DBS samples between 36 and 48 hours after birth, which is earlier than DBS sample collection in UK screening programmes (5 to 8 days after birth).[31] Minter Baerg 2018 only reported the time of DBS sample collection for the single newborn diagnosed with MPS I, which was at 58 hours after birth, limiting applicability to newborn DBS screening in the UK for these 2 studies.[31, 35] Time of DBS sample collection was unclear in the Hopkins 2015 and Chuang 2018 screening programmes.[25, 36] It is likely that this was within 72 hours for Chuang 2018, as the samples were collected as part of the Taiwanese screening programme, and therefore aligns with the UK DBS screening. For Hopkins 2015, samples were collected as part of the Missouri routine screening programme for which DBS samples are usually retrieved between 24 and 48 hours after birth, which is substantially earlier than in the UK. However, as the time of sample collection in these 2 studies is not actually reported, the applicability of both studies is unclear. The applicability of these results to UK screening is therefore unclear when timing of DBS sample collection for newborn screening is considered.

Reference standard

Only one study was at high risk of bias for this domain.[25] It was unclear whether the reference standard results were interpreted without knowledge of the index test results; as all 4 screening studies sent screen-positive samples for confirmatory testing, it is likely that this is standard procedure in diagnostic testing for MPS I, and therefore the laboratory staff were likely aware of the positive screening result. Since MPS I was confirmed using genetic analyses, this is not expected to have a large impact on the risk of bias.

While all reference standards are considered to classify the MPS I diagnosis correctly, only 1 screening study reported the use of a reference standard that diagnosed cases and the predicted phenotype, as well as cases of pseudodeficiency and carriers.[31] Two other studies distinguished confirmed cases from carriers and pseudodeficiency, but did not distinguish between the phenotypes of MPS I (for example by mutational analysis or clinical follow-up), leading to concern about applicability.[25, 35] By contrast, Chuang 2018 did not report on pseudodeficiency or carriers for MPS I, but did report that the 4 newborns diagnosed with MPS I were presumed to be affected by attenuated forms, due to being asymptomatic at follow-up (clinical evaluations were performed every 6 months following screening).

Participant flow

All studies were at high risk of bias for the participant flow domain; only newborns who screened positive using the initial index test were re-tested and received the reference standard. Repeat testing and/or use of a second-tier index test was utilised to validate the first test result and to exclude any false-positive cases. It is possible that false-negative cases, which could arise in newborns with milder forms of MPS I or from laboratory errors, may have been omitted if they were not selected for repeat testing at the initial screening stage. All or most of the screen-positive patients were included in the analysis in all 4 studies.[25, 31, 35, 36]

In the Hopkins 2015 study, it was unclear if patients received the same reference standard due to poor reporting of confirmatory testing, as it was only mentioned that newborn samples were sent to 'genetic referral centres'.[25] This study is therefore at high risk of bias.

Results

The results of screening test accuracy for MPS I are presented in Table 5. Full study details are provided in Appendix 3.

Table 5. Measures of test accuracy and test results for screening for MPS I

Study	Population	Index test	Cut-off	Reference standard	Incidence	PPV, %	FPR, %	Pseudodeficiency and Carriers
Burlina 2018[31] *Italy*	44,411 newborns	DBS IDUA activity (tandem mass spectrometry)	0.2 MoM (µmol/h)	Urinary GAG analyses IDUA testing Mutational analyses	1/44,411	7.7	NR	Pseudodeficiency: 5 Carriers: 2
Chuang 2018[36] *Taiwan*	294,196 newborns	DBS IDUA activity (tandem mass spectrometry)	<3.0 µmol/L	Urinary GAG analyses, leukocyte enzyme assay for β-IDUA and molecular DNA analysis	4/294,196 1.35 per 100,000 births	26.7	NR	Not reported
Minter Baerg 2018[35] *United States*	55,161 newborns	DBS IDUA activity (tandem mass spectrometry)	NR	Genotyping (Sanger sequencing) of IDUA gene	1/55,161	50.0	0.0018	1, but unclear whether a carrier or case of pseudodeficiency
Hopkins 2015[25] *United States*	43,701 newborns	DBS IDUA activity (fluorometric enzymatic assay)	<4.0 µmol/L/hour (0.8 percentile)	NR	3/43,701 1:14,567	11.0	0.037[a]	Pseudodeficiency: 7 Carriers: 2

Abbreviations: DBS, dried blood spot; GAG, glycosaminoglycans; IDUA, α-L-iduronidase; FPR, false positive rate; NR, not reported; PPV, positive predictive value

[a] Reported as 6-month FPR

All screening programmes detected at least one newborn with confirmed MPS I. Only one study reported on the predicted phenotype of MPS I based on the screening results; Burlina 2018 confirmed the diagnosis of the single detected case of MPS I according to high levels of urinary GAGs and a mutation that had previously been reported in a patient with Hurler-Scheie syndrome (p.Pro533Arg).[31] While the other screening studies reported the use of genetic analyses as part of the confirmatory process, none reported a predicted phenotype based on specific mutations.

Two screening studies reported on the number of newborns with a pseudodeficiency, of which there was a relatively high incidence, with 5/44,411 newborns identified in the Burlina 2018 screening study and 7/43,701 newborns in the Hopkins 2015 study. This was diagnosed upon the presence of low enzymatic IDUA activity despite normal levels of metabolites, and further confirmed by the presence of pseudodeficiency alleles using molecular DNA analysis in the Burlina 2018 study,[31] however the methods of confirming MPS I diagnoses were not adequately reported for Hopkins 2015.[25] In contrast, Minter-Baerg 2018 included pseudodeficiency under ‘false-positive’ cases along with heterozygous carriers.[35] For the single false-positive case identified, it was not specified whether this was a carrier or pseudodeficiency.

Regarding the evaluation of screening test performance, all included studies only reported results for screen-positive newborns, therefore measures of sensitivity or specificity were not reported and cannot be calculated. The only measure of test accuracy that was reported by the eligible studies was the PPV, which reflects the probability that newborns with a positive screening test truly have MPS I. However, as predictive values depend on the prevalence of the disease in the study population, it is difficult to draw comparisons across different studies. For the 3 studies which evaluated tandem mass spectrometry as a screening test for MPS I, the PPV were 7.7%, 26.7% and 50.0%, whilst the fluorometric enzyme assay achieved a PPV of 11%. It is difficult to evaluate the clinical validity of a screening test based on PPV alone, as for rare diseases such as MPS I, the PPV will be low even if the test has high sensitivity and specificity. This is further complicated by the use of different thresholds for classification of a screen-positive result, in addition to differences in the approach taken for repeat screening and use of second-tier tests with additional markers (such as with lysophosphatidylcholines and GAGs in the Minter-Baerg 2018 study) before confirmation with the reference standard.

Summary of Findings Relevant to Criterion 4: Not met

Quantity: This review identified a very limited volume of evidence on newborn screening for MPS I; only 4 studies reporting on unique newborn screening programmes were eligible for inclusion. All studies screened a large number of newborn samples (ranging from 44,411 to 294,196).
Quality: All 4 studies were of a prospective study design in large, unselected cohorts of newborn samples, and were therefore at a reduced risk of selection bias and confounding. Overall, the quality of the included studies was low; while there was a low risk of selection bias in all studies, all were at a high risk of bias in participant flow, due to the diagnosis only being confirmed in newborns who screened positive.
Applicability: All included screening studies were conducted in large, unselected newborn populations, and 3 studies were conducted in high-income countries that are considered to be reflective of the UK clinical setting. However, where reported, newborn DBS samples were collected earlier (38–58 hours after birth) than is standard in UK DBS screening (5 to 8 days after birth), limiting the applicability of results to a potential UK newborn screening programme.
Consistency: All studies screened for MPS I based on IDUA enzymatic activity, measured by tandem mass spectrometry in 3 studies, and using a fluorometric assay in combination with a pattern recognition software in the remaining study. However, there was substantial heterogeneity in the cut-offs used for classification of a screen-positive test result, the use of second-tier index tests to further rule out false-positive cases, and in the method of confirming screen-positive results. The only measure of test accuracy reported by each study was PPV, which cannot be easily compared across studies due to the influence of individual study populations. These factors together critically limit comparison of test accuracy across the different studies due to the influence of prevalence within each study population.
Conclusions: Based on the evidence assessed by this review, there is limited evidence to support that newborn DBS screening tests for MPS I are sufficiently accurate for use in a national screening programme. A small number of screening studies were identified (N=4), in which only screen-positive cases received the reference standard, increasing the risk of bias and limiting the reporting of test accuracy parameters. In conclusion, due to a limited volume of evidence, with substantial heterogeneity in screening test methods and lack of reported measures of test performance, a screening programme using these tests cannot be recommended based on the current evidence. This conclusion is consistent with that of the previous UK NSC review (2015), which also found a very limited evidence base with only 3 relevant studies included. Though further evidence on the sensitivity and specificity of tests to detect MPS I would be ideal, it is acknowledged that assessment of these test accuracy parameters is difficult to achieve in studies of screening for rare diseases. Finding mechanisms to address that is important, particularly

given the potential for identification of carriers and pseudodeficiency in MPS I screening. Further screening studies with improved methodological consistency (in terms of index test cut-offs, repeat testing and the reference standard used) may be achievable and would allow for an informative evaluation of a putative test to be used in screening for MPS I in newborn babies.

Criterion 9 — Does early initiation of treatment with HSCT and/or ERT following screening provide better outcomes compared to usual clinical care?

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

Question 2 – *Does early initiation of treatment with HSCT and/or ERT following screening provide better outcomes compared to usual clinical care?*

The 2015 UK NSC review examined the evidence related to a) treatment outcomes in screen-detected patients or patients identified pre-symptomatically through cascade testing, compared to patients detected clinically, or b) the impact of early compared with late treatment or age at treatment initiation on treatment outcomes. No studies comparing outcomes in screen- versus clinically-detected MPS I patients were identified, and 6 studies that examined early versus late treatment were identified. The conclusion of the review was that insufficient evidence, in terms of quantity and quality, was identified in the literature to draw conclusions on the impact of age at treatment on outcomes in MPS I patients.

The aim of this question was to identify and synthesise evidence published since 2014 on treatment outcomes for early initiation of treatment with HSCT and/or ERT following screening (i.e. universal newborn screening and cascade testing of siblings) compared to late initiation of treatment or usual clinical care. Studies in which the effect of age at treatment initiation on treatment outcome was explored were also eligible for inclusion.

Eligibility for inclusion in the review

This review searched for RCTs, cohort studies and interventional studies, or SLRs or MAs of these, reporting on HSCT and/or ERT in MPS I patients. Due to the small volume of evidence identified, case-control studies were also eligible for inclusion. Studies were eligible if they evaluated treatment outcomes of HSCT and/or ERT initiated early compared to late initiation or usual clinical care in patients with MPS I, or if the effect of age at treatment initiation on treatment outcome was explored in the analyses. The eligible population were patients diagnosed with MPS I.

Full details of the eligibility criteria are presented in Table 3.

Description of the evidence

A total of 13 publications were included in the review for Criterion 9. One study centre was a source of patients in 2 of the studies, it is possible that a small sample of patients from this centre was included in both cohorts, however, as the reporting of the patient sources does not allow for this to be confirmed, the studies were treated as being on independent cohorts.[37, 38] No SLRs or MAs which aligned with the scope of this review closely enough to be included were identified, and no additional relevant articles were identified through hand-searching the reference lists of the relevant SLRs.

Four studies were conducted in the US,[39-42] 2 in the UK,[43, 44] 1 in the Netherlands,[45] and 2 in both the UK and the Netherlands.[37, 46] The remaining 4 studies were international including MPS I patients from the UK, France, Germany, the Netherlands, Brazil and the US.[38, 47-49]

The majority (N=10) of studies were retrospective; 3 publications reported on prospective observational studies.[37, 38, 46] For 2 of these studies, there may have been crossover in 2 study centres, and therefore it is possible that a small sample of patients were included in both studies.[37] Eleven studies evaluated treatment outcomes of HSCT in patients with Hurler syndrome, whereas only 1 study focussed on ERT, in Hurler or attenuated MPS I patients. The remaining study reported on the effect of age at treatment initiation for HSCT and ERT together.[46] Eleven studies reported on patients with Hurler syndrome only, 1 study reported on patients with attenuated MPS I,[44] and 1 study included a mixed population (relevant results were reported for Hurler patients only).[46] The median age at treatment for Hurler syndrome ranged from 15.6 months to 21.8 months. Most studies either categorised patients into age groups or specified an age ‘cut-off’ for comparison of early and later treatment. This was often based on the median age at treatment within the study cohort, although a number of studies used a different age cut-off, ranging from 18 months to 3 years for Hurler patients.[43, 47] The reason for this varied; Javed 2018 compared outcomes for HSCT before or after 18 months based on prior guidance in the literature,[43] whereas Eisengart 2018 restricted results to 3 years in a sensitivity analysis to improve comparability of the intervention groups in the study.[47] For the 2 studies that included patients with attenuated MPS I, the median age at ERT was 5 to 10.3 years.[44, 46] Four studies examined the effect of age at treatment as a continuous variable on different treatment outcomes, and therefore only provide a relative measure for ‘earlier’ rather than ‘later’ treatment within each study cohort, limiting comparison of results in this review.

Discussion of findings

Quality assessment

The quality of the included studies was appraised using an adapted ROBINS-I checklist (Table 19; Appendix 3). A summary is presented in Table 6 and Table 7, and the full appraisal is presented in Table 20 (Appendix 3).

Table 6. Summary of ROBINS-I assessments of studies on early versus late treatment for MPS I

Question	Aldenhoven 2015a[38]	Eisengart 2018[47]	Javed 2018[43]	Laraway 2016[44]	Poe 2014[42]	Rodgers 2017[40]	Wadhwa 2019[41]	Wyffels 2017[39]
Risk of bias in confounding	Serious	Serious	Critical	Critical	Moderate	Moderate	Critical	Serious
Risk of bias in participant selection	Serious	Low	Serious	Serious	Low	Low	Serious	Critical
Risk of bias in the classification of interventions	Serious	Moderate	Moderate	Moderate	Moderate	Low	Low	Serious
Risk of bias due to deviations from intended interventions	Low	Low	Low	Low	Low	Low	Low	Low
Risk of bias due to missing data	Serious	Moderate	Serious	Low	Low	Low	Low	Serious
Risk of bias in measurement of outcomes	Serious	Low	Serious	Serious	Low	Low	Moderate	Low
Risk of bias in selection of the reported result	Serious	Serious	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate
OVERALL BIAS	SERIOUS	SERIOUS	CRITICAL	CRITICAL	MODERATE	MODERATE	CRITICAL	CRITICAL

Table 7. Summary of ROBINS-I assessments of studies which explore the effect of age at treatment as a covariate on treatment outcomes

Question	Aldenhoven 2015b[37]	Kunin-Batson 2016[48]	Langereis 2016[49]	Megens 2014[45]	Pal 2015[46]
Risk of bias in confounding	Low	Moderate	Serious	Critical	Serious
Risk of bias in participant selection	Low	Serious	Serious	Serious	Serious
Risk of bias in the classification of interventions	Serious	Serious	Serious	Serious	Serious
Risk of bias due to deviations from intended interventions	Low	Low	Low	Low	Low

Risk of bias due to missing data	**Low**	**Serious**	**Low**	**Moderate**	**Low**
Risk of bias in measurement of outcomes	**Low**	**Moderate**	**Low**	**Low**	**Low**
Risk of bias in selection of the reported result	**Moderate**	**Moderate**	**Moderate**	**Moderate**	**Moderate**
OVERALL BIAS	**MODERATE**	**SERIOUS**	**SERIOUS**	**CRITICAL**	**SERIOUS**

Confounding

The judgements for risk of bias in confounding ranged from moderate to critical, with several potential confounding baseline characteristics e.g. IDUA level, age at diagnosis, regularly not being controlled for. For 8 studies comparing outcomes for early and late treatment groups, 3 were determined to be at critical risk of bias,[41, 43, 44] with 3 at serious risk,[38, 39, 47] and 2 at moderate risk of bias.[40, 42] For 5 studies exploring the effect of age at treatment as a covariate on treatment outcomes, 1 was at critical risk,[45] 2 were at serious risk,[46, 49] 1 was at moderate risk,[48] and 1 was at low risk of bias.[37]

Participant selection

Risk of bias in participant selection was generally either serious or low, dependant on whether selection of participants was related to the outcome or the intervention. For example, in many cases participants were only included if they had survived for a certain length of time after treatment. Wyffels 2017 was judged to be at critical risk of bias for this domain because selection into the study was based on survival for at least a year and was therefore strongly related to the efficacy of treatment.[39] Of the other studies that compared early and late treatment groups, 4 were at serious risk of bias,[38, 41, 43, 44] and 3 were at low risk of bias.[40, 42, 47] For studies looking at effect of age at treatment as a covariate on clinical outcomes, 4 were at serious risk of bias,[45, 46, 48, 49] and 1 was at low risk.[37]

Classification of interventions

The early and late groups for treatment initiation were generally well-defined; as such, 2 studies comparing early and late treatment were judged to be at a low risk of bias,[40, 41] 4 at moderate risk of bias,[42-44, 47] with only 2 determined to be at a serious risk of bias.[38, 39] However, for the group of studies exploring age at treatment as a covariate on MPS I outcomes, given that this was investigated as a continuous variable (and therefore intervention groups were not clearly defined), all 5 studies were found to be at a serious risk of bias.[37, 45, 46, 48, 49]

Deviation from intended interventions

All studies were found to be at low risk of bias due to deviations from intended interventions. This was because many focussed on HSCT, which is typically a standalone treatment that is given by intravenous infusion and therefore cannot be discontinued. When HSCT and ERT treatments were

both investigated, the study was judged to be at a low risk of bias as deviating between interventions by switching between HSCT and ERT is expected in standard clinical practice.

Missing data

There was variability in if and how studies reported exclusion of participants due to missing data and for 5 this was not clear. This affected the judgement of the risk of bias, with 4 studies at serious risk,[38, 39, 43, 48] 2 at moderate risk,[45, 47] and 7 at a low risk of bias due to missing data.[37, 40-42, 44, 46, 49]

Measurement of outcomes

As most interventional studies were retrospective chart reviews, the individual assessing the outcomes was typically the patient's own clinician and would have therefore been aware of the intervention received. However, whether outcome measurements could be influenced by knowledge of the intervention received, varied, and as such the risk of bias was higher for some studies than for others. For studies comparing early and late HSCT, 4 were determined to be at low risk of bias,[39, 40, 42, 47] 1 at moderate risk,[41] and 3 at serious risk of bias in measurement of outcomes.[38, 43, 44] Amongst studies exploring the effect of age at treatment as a covariate on MPS I outcomes, 4 were judged to be at low risk,[37, 45, 46, 49] and 1 at moderate risk of bias.[48]

Selection of the report result

As no studies provided clear evidence that all results were reported, the risk of bias in selection of the reported result was moderate or above. For studies comparing early and late treatment groups, 2 were judged to have a serious risk of bias due to selective reporting of patient subgroups or only clinical outcomes that changed significantly,[38, 47] whilst the remaining 6 were moderate.[39-44] All 5 studies that evaluated age at treatment as a covariate on outcomes were judged to be at a moderate risk of bias.[37, 45, 46, 48, 49]

Overall bias

For the studies comparing early and late treatment groups, Javed 2018, Laraway 2016, Wadhwa 2019 and Wyffels 2017 were judged to be at critical risk of bias and Aldenhoven 2015a and Eisengart 2018, at serious risk of bias. Only Rodgers 2017 and Poe 2014 were at moderate overall risk of bias. For studies looking at age at treatment as a covariate on MPS I outcomes, Aldenhoven 2015b was judged to be at a moderate risk of overall bias, Kunin-Batson 2016, Langereis 2016 and Pal 2015 were at serious risk of overall bias and Megens 2014 was at critical risk of overall bias.

Results

The results of the included studies are presented in Table 8. Full study details are provided in the study-level data in Appendix 3.

Table 8. The association between age at treatment initiation and clinical outcomes for MPS I

<table>
<tr><th>Study</th><th>Population</th><th>Treatment</th><th>Age at treatment</th><th>Age threshold or groups for analysis</th><th>Treatment outcomes (early versus late initiation)</th></tr>
<tr><td>Aldenhoven 2015a[38]</td><td>Hurler syndrome (N=217)</td><td>HSCT</td><td>Median 16 months (range 2–47 months)</td><td>Continuous (primary outcome)

16 months (secondary outcomes)</td><td>Primary outcome – neurodevelopment (measured by DQ/IQ)
Later HSCT was a significant predictor of inferior DQ/IQ (β: −8.40; 95% CI: −14.62 to −2.19; p=0.009).

Secondary outcomes
Later HSCT was associated with poorer outcomes for neurological and cardiac endpoints and carpal tunnel syndrome compared with HSCT before 16 months of age.
<table><tr><th></th><th colspan="2">%</th><th>OR/HR (95% CI)[a]</th><th>p-value</th></tr><tr><th></th><th><16 months</th><th>>16 months</th><th></th><th></th></tr><tr><td>Cerebral atrophy</td><td>23</td><td>46</td><td>OR 3.22 (1.60–6.50)</td><td>0.001</td></tr><tr><td>Cord compression</td><td>5</td><td>16</td><td>HR 2.84 (1.02–1.41)</td><td>0.04</td></tr><tr><td>Carpal tunnel syndrome</td><td>33</td><td>56</td><td>HR 1.72 (1.11–2.68)</td><td>0.02</td></tr><tr><td>Mitral valve insufficiency</td><td>26</td><td>47</td><td>OR 2.46 (1.30–4.65)</td><td>0.006</td></tr><tr><td>Atrial valve insufficiency</td><td>19</td><td>37</td><td>OR 2.40 (1.19–4.82)</td><td>0.01</td></tr></table>[a] <16 months was the reference, with an OR of 1</td></tr>
<tr><td>Eisengart 2018[47]</td><td>Hurler syndrome (N=93)</td><td>ERT (intervention group)

HSCT or no treatment (control groups)</td><td>Median age 1.3 years (range 0.5–2.7)</td><td>Continuous 3 years</td><td>Survival
Differences in survival between the untreated group and the ERT group, and between the HSCT group and ERT group remained qualitatively similar when age at treatment was restricted to <3 years (HR=2.4, p=0.046 and HR=2.50; p=0.089, respectively).

The median age of death did not change significantly if ERT was initiated prior to 3 years of age compared to the overall cohort (8.9 vs. 9.0 years)

CNS outcomes
Restricting age at treatment to <3 years did not significantly change CNS outcomes</td></tr>
<tr><td>Javed 2018[43]</td><td>Hurler syndrome (N=26)</td><td>HSCT</td><td>Mean 12.7 months (range 4–24)</td><td>18 months</td><td>DS/CS ratios</td></tr>
</table>

<table>
<tr><td></td><td></td><td></td><td></td><td></td><td>DS/CS ratios which were significantly higher in patients with severe corneal clouding (p=0.043), were also significantly higher in patients treated under 18 months (p=0.023)

Iduronidase levels
No association between age at transplant and iduronidase enzyme levels, which in those treated <18 months, remained significantly lower in the severe versus mild/moderate corneal clouding groups (p=0.02)</td></tr>
<tr><td>Poe 2014[42]</td><td>Hurler syndrome (N=31)</td><td>HSCT</td><td>Median 13.8 months (range 2.1–34.3)</td><td>Continuous

Group 1: 2–8 months, n=6

Group 2: 9–18 months, n=17

Group 3: ≥18 months, n=8</td><td>Cognitive function
• Earlier transplantation was associated with greater gains in cognitive function (β: −0.024, p<0.001)

Adaptive behaviour
• Earlier transplantation was associated with better post-transplant adaptive behaviour development (β: −0.013, p=0.030)

Language skills
• Earlier transplantation was associated with better skill development (receptive language, β: −0.022, p=0.004; expressive language, β: −0.023, p=0.0010)

Audiological and visual function
• Hearing loss was not associated with age at transplantation
• Visual function did not differ across the 3 age groups</td></tr>
<tr><td>Rodgers 2017[40]</td><td>Hurler syndrome (N=134)</td><td>HSCT</td><td>Mean 21.8 months (SD 20.8)</td><td><12 months
12 to 24 months
>24 months</td><td>Age at HSCT was not significantly associated with survival over the first 8 years post-HSCT:

<table><tr><th>Covariate</th><th>Hazard ratio</th><th>95% CI</th><th>p-value</th></tr><tr><td>Age at treatment <12 month</td><td>Ref</td><td>-</td><td>-</td></tr><tr><td>Age at treatment 12–24 months</td><td>1.42</td><td>0.59 to 3.40</td><td>0.429</td></tr><tr><td>Age at treatment >24 months</td><td>1.45</td><td>0.52 to 4.00</td><td>0.475</td></tr></table></td></tr>
<tr><td>Wadhwa 2019[41]</td><td>Hurler syndrome (N=96)</td><td>HSCT</td><td>Median 1.5 years (range 0.4–6.0)</td><td>1.5 years</td><td>Early HSCT was associated with a lower risk of all-cause late mortality (HR 0.2; 95% CI: 0.05 to 0.9; p=0.03)</td></tr>
<tr><td>Wyffels 2017[39]</td><td>Hurler syndrome (N=74)</td><td>HSCT

HSCT and ERT</td><td>Group 1: Mean 1.8 years (range 0.5–6.0)

Group 2: Mean 1.5 years (0.4–2.9)</td><td>2 years</td><td>No significant difference in the incidence of carpal tunnel syndrome between early (48%; 95% CI, 32–62) and late (47%; 95% CI 21–68) HSCT treatment groups</td></tr>
</table>

Laraway 2016[44]	Patients with attenuated MPS I (N=35)	ERT (laronidase)	Median 11.3 years (range 0.5–23.1)	10 years	**Mitral valve deterioration** Fewer children aged <10 years at treatment initiation experienced deterioration compared with patients aged ≥10 years (14% vs 45% at the last assessment) **Aortic valve** Fewer children aged <10 years at treatment initiation experienced aortic valve deterioration compared with patients aged 10 years (14% vs 40%) **Corneal clouding** Fewer children aged <10 years at treatment initiation experienced corneal clouding deterioration than patients aged ≥10 years (9% vs 25%) **Visual acuity** A greater percentage of children aged <10 years at treatment initiation deteriorated compared with patients aged ≥10 years (40% vs 14%)
Association between treatment outcomes and age at treatment initiation as a continuous measure only					
Aldenhoven 2015b[37]	Hurler syndrome (N=56)	HSCT	Median 13.5 months (range 3–44)	Continuous	HSCT at a later age was a predictor for: • acute GVHD: HR 1.13 (95% CI 1.05–1.21); p=0.001 • chronic GVHD: HR 1.08 (95% CI 1.02–1.15); p=0.01 • CMV reactivation: HR 1.09 (95% CI 1.01–1.18); p=0.02
Kunin-Batson[48]	Hurler syndrome (N=47)	HSCT	Mean 18.5 months (SD 8.2)	Continuous	**Cognitive and adaptive functioning** Age at transplant was not a significant predictor of cognitive (IQ) or adaptive functioning **Physical functioning** Later age at transplant was significantly associated with poorer physical functioning (β*:−8.10, 95% CI −13.16 to −3.05, p=0.002).
Langereis 2016[49]	Hurler syndrome (N=52)	HSCT	Median 12 months (range 3–30)	Continuous	No significant interaction between radiographic parameters (for hip dysplasia) and age at transplantation
Megens 2014[45]	Hurler syndrome (N=17)	ERT prior to HSCT	ERT: Median 14 months (range 7–43) HSCT: Median 18 months (10–43)	Continuous	No association between the incidence of perioperative airway management difficulty and age at treatment initiation (OR 1.01, p=0.36)
Pal 2015[46]	Patients with MPS I (Hurler syndrome [N=44] and attenuated [N=17])	HSCT (Hurler) ERT (Hurler and attenuated patients)	Median 18 months (range 3–364)	Continuous	Later start of treatment significantly correlated with the need for therapeutic airway intervention following initiation of HSCT/ERT (p=0.012)

Where given, the beta coefficient (β) measures the degree of change in an outcome variable for every unit of change in the predictor variable. When the beta coefficient is significant (as determined by the p-value), a positive value indicates that for every unit increase in the predictor variable, the outcome

variable will increase by the beta coefficient value. A negative value indicates that for every unit increase in the predictor variable, the outcome variable will decrease by the beta coefficient value. *Measure not specified but assumed to be the β coefficient based on study-reported statistical methodology.
Abbreviations: CI: confidence interval; CMV: cytomegalovirus; DS/CS: dermatan sulphate/chondroitin sulphate; DQ/IQ: developmental quotient/intelligence quotient; ERT: enzyme replacement therapy; GVHD: graft-versus-host disease; HR: hazard ratio; HSCT: haematopoietic stem cell transplantation; MPS I: mucopolysaccharidosis type I; OR: odds ratio.

HSCT

The included interventional studies reported varied clinical outcomes on which early and later HSCT (+/- ERT) could be compared (Table 8). The majority investigated Hurler syndrome patients. Whilst Rodgers 2017 found that age at HSCT was not significantly associated with survival (<12 months vs 12–24 months: p=0.429; <12 months vs >24 months: p=0.475),[40] Wadhwa 2019 demonstrated that early BMT was associated with lower risk of all-cause late mortality (p=0.03) thereby concluding that younger age at BMT was protective in these patients with Hurler syndrome, although the risk of all-cause early mortality was not investigated by this study.[41] Another study looking at cardiac outcomes found that earlier HSCT was significantly associated with less mitral (p=0.006) and atrial valve insufficiency (p=0.01).[38] Aldenhoven 2015a and Poe 2014 both demonstrated that earlier HSCT benefits neurodevelopmental and cognitive outcomes, including cerebral atrophy (p=0.001), cord compression (p=0.04), cognitive function (p<0.001), adaptive behaviour (p=0.03) and language skills (p=0.01).[38, 42] In contrast, Kunin-Batson 2016 found that age at transplant was not a significant predictor or cognitive (IQ) or adaptive functioning.[48] The effect of carpal tunnel syndrome was also examined, with Aldenhoven 2015a finding that carpal tunnel syndrome was less common with earlier treatment (p=0.02),[38] but Wyffels 2017 finding no significant difference in carpal tunnel syndrome incidence between the early and late treatment groups.[39] A study examining biomarkers of transplantation established that differences in dermatan sulphate/chondroitin sulphate ratio and iduronidase enzyme levels associated with severe corneal clouding, are not related to age at HSCT.[43] Thus, while some studies suggest that earlier initiation of HSCT is beneficial for some clinical outcomes, there are also others where no difference between early and late treatment initiation has been found. This may be due to significant heterogeneity in the patient population as well as data collection methods (e.g. different age at which groups were split into earlier and late).

ERT

Only 2 studies examined how age at ERT treatment affects clinical outcomes, with 1 evaluating this in Hurler syndrome patients specifically and the other in attenuated MPS I (Table 8). Eisengart 2018 reported that in Hurler syndrome patients, age at death did not significantly change when evaluating the group of patients for whom ERT was initiated prior to 3 years of age, compared to the overall cohort.[47] Laraway 2016 investigated cardiac outcomes, corneal clouding, and visual acuity in patients with attenuated MPS I, finding that fewer children aged <10 years at ERT experienced mitral/aortic valve deterioration and corneal clouding, but a greater proportion presented with reduced visual acuity.[44] However, this study did not conduct statistical analyses and, as the participants studied had an attenuated MPS I phenotype, the results are not comparable with the other studies.

<u>Age at treatment initiation as a continuous variable</u>

Five studies investigated the effect of age at treatment initiation on clinical outcomes as a continuous measure only. Amongst studies evaluating Hurler syndrome patients, later age at HSCT was shown to be associated with poorer physical functioning (p=0.002),[48] and a predictor for acute/chronic graft-versus-host disease (p=0.001 and p=0.01, respectively) and cytomegaly virus reactivation (p=0.02).[37] Further, while Megens 2014 found no association between the incidence of airway management difficulty and age at ERT+HSCT treatment initiation (p=0.36),[45] Pal 2015 demonstrated a greater requirement for therapeutic airway intervention after treatment initiation with ERT or HSCT (p=0.012).[46] No significant associations between age at HSCT and radiographic parameters for hip dysplasia were found.[49]

Conclusions

No effect of age of treatment initiation on certain outcomes was demonstrated in some studies, though others did report an association between age at initiation of HSCT or ERT and clinical outcomes for MPS I patients. However, in many studies the size of the effect, although statistically significant, is relatively small and it is therefore unclear whether early diagnosis of MPS I would result in a clinically significant improvement in patients' symptoms. Furthermore, the age by which treatment must be started to experience any benefits cannot be determined as treatment groups with specific thresholds have not been not compared.

The majority (11/13) of studies focussed solely on Hurler patients, whilst the effect of early initiation of treatment for patients with attenuated MPS I was rarely investigated. Therefore, for these patients the importance of the newborn screening test, which would detect attenuated forms of MPS I, is not well characterised.

Another limitation is that studies generally did not evaluate infants who had initiated treatment at an age where newborn screening would be valuable; the median age at treatment across all studies was more than 1 year and although infants with age at treatment as low as 2 months were included, there were insufficient numbers of these patients for whom data was reported individually to draw conclusions relating to the value of MPS I detection through newborn screening. This is consistent with the previous (2015) UK NSC review, which noted that the median age of treatment in studies was more aligned with clinical detection than earlier initiation of treatment following detection through screening. Overall, there is insufficient evidence to determine whether early initiation of HSCT or ERT improves clinical outcomes for MPS I patients. Generating research evidence of early treatment effectiveness is challenging for rare diseases. However, evaluation of outcomes of 'early' treatment in sibling cases may assist in building the evidence base. Furthermore, improved consistency in the investigated treatment outcomes would be helpful to minimise the uncertainties identified in this review.

Summary of Findings Relevant to Criterion 9: Not met

Quantity: The volume of evidence identified in this review was relatively small; 8 studies compared treatment outcomes for early versus late initiation of HSCT and/or ERT, the majority of which exclusively recruited patients with severe MPS I Hurler syndrome phenotype. Five additional studies were identified that evaluated the association between treatment outcomes and age at treatment as a continuous measure, which also focussed on Hurler syndrome. Sample populations in 2 of the 13 studies were of moderate size (between 134 and 217 MPS I patients),[38, 40] whilst for the remaining studies population sizes were relatively small (19 to 96 MPS I patients). The median age at treatment was >1 year of age across all studies, which is consistent with clinical detection of disease, with no evidence on the effect of early initiation of treatment following screening, especially in very young infants (less than 2 months of age) or in siblings of known cases.

Quality: Ten of the total 13 studies were retrospective and were therefore at an increased risk of selection bias and confounding, with the remaining 3 of a prospective study design. The quality of the included studies was generally low, with significant risk of bias amongst studies in participant selection, where inclusion of patients was often associated with their survival or other treatment outcomes. Classification of the timing of intervention (i.e. splitting participants into age groups based on treatment initiation for inference of 'early' or late' treatment following diagnosis) was also found to be inconsistent and poorly justified in many studies, particularly where age at treatment was measured as a continuous variable. Deviation from the intended intervention was deemed to be unlikely (where HSCT cannot be discontinued), or to reflect standard clinical practice, and was therefore judged to be at low risk of bias across all studies.

Applicability: All eligible studies were conducted in high-income countries that are considered to be reflective of the UK setting. Thus, the applicability to UK clinical practice is high.

Consistency: A large variety of clinical outcomes were evaluated by the included studies. Where more than one study measured the same outcomes the methods of measurement and the classification of 'early' and 'late' treatment groups differed, meaning that informative comparisons between studies are difficult to make. Overall, some studies suggest a statistically significant benefit for earlier initiation of treatment, but the effect is small and of unknown clinical significance; other studies find no difference between early and late treatment initiation. Therefore, there is high uncertainty as to whether earlier treatment provides an overall clinical benefit for patients with MPS I.

Conclusions: Based on the evidence synthesised by this review, the question of whether early initiation of treatment improves outcomes for MPS I patients cannot be answered. The results were mixed and no clear threshold for the age at which any potential effect may exist has been established. Although there may be benefits in initiating HSCT early in patients diagnosed with Hurler syndrome for survival, neurodevelopmental, cognitive, and cardiac outcomes, these outcomes were often conflicting across the studies. In addition the evidence was heterogeneous and limited by study design, methodology, and small sample sizes. With significant heterogeneity in patient baseline characteristics and the way outcomes were assessed, results are not comparable. Studies also did not directly assess the potential benefit of early treatment following positive newborn screening tests, and the cut-offs used were often comparable to the median age of treatment for clinically detected Hurler syndrome, rather than early initiation of HSCT/ERT in a potentially screen-detected newborn. With only one included study that specifically examined the effect of age at treatment initiation in patients with attenuated forms of MPS I, the potential benefit of screening and subsequent early treatment is particularly unclear for this phenotype. In summary, based on the findings of this review, there is insufficient evidence to conclude whether newborn screening for MPS I has an impact on early treatment of newborns confirmed to have MPS I, and whether this is beneficial for these infants.

Review summary

Conclusions and implications for policy

Based on the overall synthesis of evidence published since the last UK NSC review in 2014, newborn screening for MPS I is still not recommended.

Two questions were considered in this rapid review: whether there has been a significant development in the evidence base relating to (1) an appropriate screening test for the identification of newborns with MPS I and (2) a treatment benefit of HSCT and/or ERT initiated early in MPS I patients following screening or identification by cascade-testing of siblings.

There were several limitations to the evidence. Firstly, only 4 relevant studies reporting on newborn screening for MPS I were published since the searches for the last UK NSC review were run in 2014.[25, 31, 35, 36] Three studies reported on assessing IDUA activity measured by tandem mass spectrometry,[25, 31, 36] while only 1 study assessed the use of a fluorometric assay of IDUA activity as a screening test for MPS I.[35] Differences in the screening process further limit comparability of the screening studies; there was substantial heterogeneity in the cut-offs used for classification of a screen-positive test result, along with the approach taken for repeat testing and use of second-tier index tests to further rule out false-positive cases.

Crucially, full reporting of test accuracy for assessment of IDUA activity by tandem mass spectrometry or fluorometric assay was lacking in all studies; only PPV was reported without any measure of variation such as a confidence interval, diminishing confidence in the results. While positive or negative predictive values are often considered valuable for assessing the clinical validity of a screening test, they depend on the prevalence of the condition in the population. As such, even a highly accurate screening test will have a poor PPV when used in a low-prevalence population,[50] as may be the case for MPS I. In all newborn screening programmes identified by this review, only screen-positive samples were sent for confirmatory testing. As such, the number of true or false-negative test results was not determined, preventing calculation of test accuracy parameters such as sensitivity, specificity and likelihood ratios, which are not substantially influenced by disease prevalence and are intrinsic to the screening test.[51] It is important to acknowledge that this process is often characteristic of screening for rare diseases; due to the expected low incidence of the condition, the ultimate aim is to identify all babies at risk of having the condition for follow-up and confirmatory testing, while excluding those who do not require follow-up. Repeat rounds of screening with the index test are considered to increase

confidence in excluding negative test results from the screening process without receiving the reference standard.[36] Nevertheless, the small number of studies along with study heterogeneity and lack of test accuracy parameters, all critically limit the evaluation of test accuracy for the newborn screening programmes identified in this review.

The phenotype of confirmed MPS I was specified in only 2 of the included studies.[31, 36] In clinical practice, MPS I phenotype and subsequent treatment choices are often based on clinical symptoms, which usually present later in infancy for Hurler syndrome or later in life for attenuated forms, therefore resulting in a diagnostic delay and potentially increasing the risk of disease progression, particularly for those with Hurler syndrome.[2] By undertaking genetic analyses, it is possible to predict the phenotype before the onset of symptoms if a mutation's correlation with a phenotype has been established, as determined by Burlina 2018.[31] In the wider literature, Kingma 2013 has reported on a diagnostic algorithm involving enzymatic analysis of IDUA activity in fibroblasts that allowed for differentiation between Hurler syndrome and attenuated MPS in affected newborns with 82% sensitivity and 100% specificity, although the performance of this algorithm is yet to be validated in a larger cohort.[2] No evidence on an index test that can effectively distinguish between MPS I phenotypes in newborn samples was identified by this review. For newborns confirmed to have MPS I through screening, it may not be possible to determine how severely the child will be affected by the condition until symptoms arise, which could be mild and may present much later in life if affected by an attenuated form. For example, Chuang 2018 predicted that the diagnosed newborns had attenuated forms of MPS I, as they had remained asymptomatic "to date".[36] This raises potential implications for treatment choices and could also impact the wellbeing of the parent(s) or caregivers following the screening test result, due to anxiety associated with not knowing when or how the disease will progress.

Two screening programmes explicitly reported the number of cases of pseudodeficiency for MPS I. Pseudodeficiency alleles can result in reduced enzymatic activity of IDUA *in vitro* in samples from people who do not have MPS I, thereby complicating newborn DBS screening for MPS I.[18] Pseudodeficiency is not known to lead to any disease or clinical symptoms, and therefore treatment is not required.[18] The relatively high incidence of pseudodeficiency detected in these 2 studies highlights the importance of using multiple confirmatory tests to distinguish screen-positives who are affected by MPS I and those who have reduced IDUA activity without MPS I. The prevalence of pseudodeficiency has been reported to be particularly high in African-American populations or those of African origin;[18, 22] indeed, it appears that all newborns with pseudodeficiency identified in the Burlina 2018 study were of African descent.[31] As ethnicity of the screened population were not reported by Hopkins 2015, it is not possible to examine whether this may have influenced the incidence of pseudodeficiency in this study.[25, 31] By contrast, the Minter-Baerg 2018 screening programme classified both heterozygotic carriers and newborns with

pseudodeficiency as false-positives cases.[35] Despite the use of similar confirmatory tests as performed in the Burlina 2018 screening programme, Chuang 2018 did not distinguish between confirmed MPS I cases and pseudodeficiency.[36] This is a very important factor to consider when evaluating screening programmes for MPS I, as inconsistency in the identification or definition of pseudodeficiency could further complicate screening test results and evaluation of test performance within and across screening studies.

As well as sparse evidence to support the accuracy of DBS screening tests for MPS I, the evidence was limited and of poor quality for the effect of early versus late treatment initiation for infants or children diagnosed with MPS I. All but 2 were retrospective studies of clinical data; as such, they were at a high risk of selection bias resulting from inclusion of MPS I patients with complete data only and were further limited by samples sizes ranging from 26 to 217 patients, where application of statistical methodology may not produce robust results. Furthermore, no studies specifically aimed to evaluate treatment outcomes for MPS I following diagnosis by screening or cascade testing of siblings. Instead there was substantial heterogeneity in the methodology and measured outcomes, with no clear trend towards outcome improvement with early or late treatment initiation. The median age at treatment initiation varied between studies, and the reporting of relevant results was often sparse, with 4 studies only briefly reporting whether outcomes were associated with the age of treatment initiation.

Overall, there was some evidence that, if given early (i.e. before the median age of treatment initiation or younger than the chosen cut-off), HSCT was associated with favourable neurodevelopmental and cognitive treatment outcomes,[38] lower risk of all-cause mortality,[41] and lower incidence of mitral/atrial valve insufficiency and carpal tunnel syndrome in infants with Hurler syndrome.[38] By contrast, 2 studies found no significant difference between early and late treatment for survival or incidence of carpal tunnel syndrome, respectively.[39, 40] When considered as a continuous variable, later age at HSCT showed a weak statistically significant association with poor physical function and the need for therapeutic airway intervention. [46, 48] Whilst statistically significant interactions were detected between age at treatment and graft-versus-host disease (both acute and chronic) and cytomegalovirus reactivation,[38] the relative measures of effect were very small. No associations were found between age at treatment and cognitive and adaptive functioning,[48] or hip dysplasia.[49]

Ultimately, while some studies reported a statistically significant benefit of early HSCT or ERT others did not and the overall effect sizes were small. It is difficult to form conclusions based on the current evidence due to the variety of outcomes reported and the fact that the age by which early and late treatment groups were defined varied between studies, or that age was only analysed as a covariate of the outcome measured.

Similarly, as for HSCT, the effect of age at ERT initiation on outcomes in Hurler syndrome or attenuated MPS I patients remains unclear. In Hurler patients, there was no clear difference in age at death for early ERT (before 3 years) compared to an overall cohort (treatment ranging from 0.5 to 4.7 years of age).[47] One study reported favourable treatment outcomes in attenuated MPS I patients who had received ERT before the median age of 10 years compared with after 10 years, but these findings were not supported by statistical analyses. Furthermore, these patients were not identified through newborn screening; instead diagnoses were made based on both symptoms and molecular analyses at a median of 4 years of age, suggesting that for many patients there was a large gap between diagnosis and treatment. Evidence on screening or early treatment for attenuated MPS I was particularly limited overall. Thus, the benefit of treating screen-detected patients with attenuated forms of MPS I, who typically present with symptoms later in childhood, remains unknown. Globally, over half (60.5%) of patients with MPS I are estimated to have Hurler syndrome,[3] therefore the lack of evidence for attenuated phenotypes may be unsurprising. Nevertheless, it is important to understand how newborn screening for MPS I would ultimately affect patients confirmed to have MPS I across the phenotypic spectrum. The newborn diagnosed with Hurler-Scheie syndrome following DBS testing in the Burlina 2018 screening study was reportedly treated with ERT, but details on how soon after diagnosis treatment was initiated, along with any clinical outcomes, were not reported.[31]

Limitations

This section considers limitations of the review methodology. Limitations of the evidence and evidence gaps are discussed in the section above.

This rapid review was conducted in line with the UK NSC requirements for evidence summaries, as described at https://www.gov.uk/government/publications/uk-nsc-evidence-review-process/appendix-f-requirements-for-uk-nsc-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 21 in Appendix 6).

Searches of multiple databases were conducted (see Appendix 3). Database search terms were restricted by study design and interventions and limited to studies published since 2014. Published and well validated filters were used to limit by study design,[12, 26, 34] searches were supplemented with SLR reference list searches, and expert clinical opinion was sought on the completeness of the list of relevant records identified, which decreases the likelihood that major important studies were missed.

Included publication types

This review only included peer-reviewed journal publications and excluded publications that were not peer-reviewed and grey literature. This may have led to the exclusion of relevant evidence. However, this is an accepted methodological adjustment for a rapid review and is unlikely to miss any pivotal studies.

No calculations were performed in this review. For question 1, publications were excluded if they only presented data that would require a calculation of test accuracy parameters that were otherwise not reported. This was taken as a pragmatic approach and was unlikely to result in key screening studies being missed.

Language

Only studies published in English were included. Given that this review was focusing on evidence relevant to the UK setting, this limitation should not have led to the exclusion of any pivotal studies.

Review methodology

Articles were reviewed by a single reviewer in the first instance. A second reviewer examined all included articles, 10% of excluded articles, and any articles where there was uncertainty about inclusion. Systematic reviews were identified through a separate search and were pre-screened based on title by a single, senior reviewer. This pragmatic strategy should have minimised the risk of errors.

Appendix 1 — Search strategy

Electronic databases

The search strategy included searches of the databases shown in Table 9. MEDLINE, MEDLINE In-Process, MEDLINE Daily, Epub Ahead of Print, Embase, and the Cochrane Library, including Cochrane Database of Systematic Reviews (CDSR), Cochrane Central Register of Controlled Trials (CENTRAL) and Database of Abstracts of Reviews of Effects (DARE).

Table 9. Summary of electronic database searches and dates

Database	Platform	Searched on date	Date range of search
MEDLINE, MEDLINE In-Process, MEDLINE Daily, Epub Ahead of Print	Ovid SP	2 May 2019	1946 to Present
Embase	Ovid SP	2 May 2019	1974 to 2016 July 01
The Cochrane Library, including: - Cochrane Database of Systematic Reviews (CDSR) - Cochrane Central Register of Controlled Trials (CENTRAL)	Wiley Online	2 May 2019	CDSR: Issue 7 of 12, July 2016
Database of Abstracts of Reviews of Effects (DARE)	Centre for Reviews and Dissemination, University of York	2 May 2019	DARE: Issue 2 of 5, April 2015

Search terms

Search terms included combinations of free text and subject headings (Medical Subject Headings [MeSH] for MEDLINE, and Emtree terms for Embase). Due to the small size of the evidence base, searches were based on disease area (MPS I) terms, limited only by publication type, date of publication (since the previous evidence synthesis was conducted) and to studies conducted in humans.

Search terms for MEDLINE, MEDLINE In-Process, MEDLINE Daily, Epub Ahead of Print and Embase are shown in Table 10, search terms for the Cochrane Library databases are shown in Table 11 and search terms for DARE are shown in Table 12.

Table 10. Search strategy for MEDLINE, MEDLINE In-Process, MEDLINE Daily, Epub Ahead of Print and Embase (to be searched simultaneously via the Ovid SP platform)

Term Group	#	Search Terms	Results
Disease terms	1	mucopolysaccharidosis i/ or mucopolysaccharidosis 1/	4351
	2	(((Mucopolysaccharidosis or MPS) adj2 (type 1 or type i)) or ((mucopolysaccharidosis or MPS) adj ("1" or I or IH or IS or IH-S))).ti,ab.	3179
	3	(Hurler* or Scheie* or Pfaundler-Hurler* or ((alpha-L-Iduronidase or iduronidase or IDUA) adj3 deficien*) or Gargoylism or Lipochondrodystrophy).ti,ab.	3907
	4	Iduronidase/	1687
	5	mucopolysaccharidosis/ or hurler syndrome/ or scheie syndrome/ or Hurler Scheie Syndrome/	9252
	6	levo iduronidase/	1093
	7	or/1-6	11349
Limits	8	("conference abstract" or "conference review").pt.	3402051
	9	exp animals/ not exp humans/	8999030
	10	(comment or letter or editorial or "case reports").pt.	5143325
	11	(case stud$ or case report$).ti.	594106
	12	historical article/	351212
	13	case study/	2033043
	14	or/8-13	17787284
Combined	15	7 not 14	7908
	16	limit 15 to yr="2014-current"	1414
	17	remove duplicates from 16	1027

Table 11. Search strategy for the Cochrane Library Databases (Searched via the Wiley Online platform)

Term Group	#	Search Terms	Results
Disease area	1	[mh ^"mucopolysaccharidosis 1"]	17
	2	(((Mucopolysaccharidosis or MPS) NEAR/2 (type 1 or type i)) or ((mucopolysaccharidosis or MPS) NEAR/1 ("1" or I or IH or IS or IH-S))):ti,ab,kw	209
	3	(Hurler* or Scheie* or Pfaundler-Hurler* or ((alpha-L-Iduronidase or iduronidase or IDUA) NEAR/3 deficien*) or Gargoylism or Lipochondrodystrophy):ti,ab,kw	47
	4	[mh ^iduronidase]	6
	5	[mh ^mucopolysaccharidosis] or [mh ^"Hurler Syndrome"] or [mh ^"Scheie Syndrome"] or [mh ^"Hurler Scheie Syndrome"]	29
	6	[mh ^"levo iduronidase"]	0
	7	{OR #1-#6}	248
Limits	8	("Conference abstract" or "conference review"):pt	145879
	9	#7 NOT #8	204
Total	10	#9 in Cochrane Reviews and Cochrane Protocols with Cochrane Library publication date from Jan 2014 to Mar 2019	6
		#9 in Trials with Publication Year from 2014 to 2019	64

Table 12. Search strategy for DARE (Searched via the CRD website)

Term Group	#	Search Terms	Results
Disease terms	1	MeSH DESCRIPTOR Mucopolysaccharidosis I IN DARE	1
	2	(((Mucopolysaccharidosis or MPS) NEAR2 (type 1 or type i)) or ((mucopolysaccharidosis or MPS) NEAR1 ("1" or I or IH or IS or IH-S))) IN DARE	4
	3	(Hurler* or Scheie* or Pfaundler-Hurler* or ((alpha-L-Iduronidase or iduronidase or IDUA) NEAR3 deficien*) or Gargoylism or Lipochondrodystrophy) IN DARE	2
	4	MESH DESCRIPTOR iduronidase IN DARE	0
	5	(MeSH DESCRIPTOR mucopolysaccharidosis or MESH DESCRIPTOR Hurler Syndrome or MESH DESCRIPTOR Scheie Syndrome or MESH DESCRIPTOR Hurler Scheie Syndrome) IN DARE	0
	6	MeSH DESCRIPTOR "levo iduronidase" IN DARE	0
	7	#1 OR #2 OR #3 OR #4 OR #5 OR #6	5
	8	*IN DARE FROM 2014 TO 2019	9540
	9	#7 and #8	1

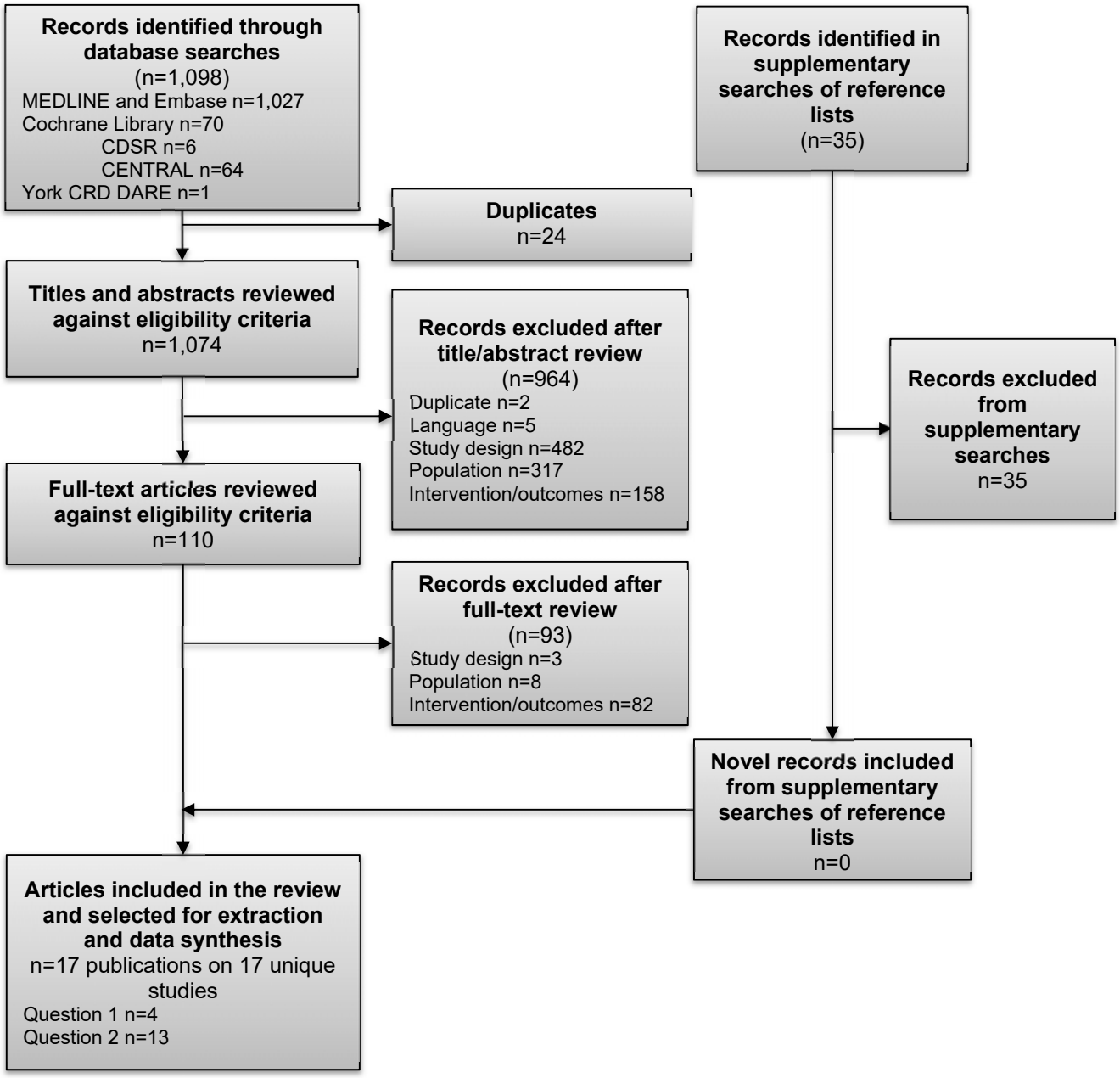
Results were imported into EndNote and de-duplicated.

Appendix 2 — Included and excluded studies

PRISMA flowchart

Figure 3 summarises the volume of publications included and excluded at each stage of the review. Seventeen 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 3. Summary of publications included and excluded at each stage of the review



Publications included after review of full-text articles

The 17 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 following approach would be taken to prioritise studies for extraction:

1. Systematic reviews and meta-analyses would be considered the highest quality of evidence if any were found. Following this, study designs would be prioritised for each question in the order listed in Table 2 and Table 3 respectively.
2. Studies relating to epidemiology would be prioritised if they considered a UK population, followed by studies from Western populations analogous to the UK.

In addition, the following criteria were applied after assessing the overall volume of evidence identified in the review:

3. Epidemiology studies that were completed over 10 years before this review was conducted (i.e. studies that were completed in 2005 or earlier, regardless of publication date) were not extracted.

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

Table 13. Summary of publications included after review of full-text articles, and the question(s) each publication was identified as being relevant to

Study	Question	Screening test (Q1)/intervention (Q2)
Burlina 2018[31]	Q1	DBS IDUA activity (tandem mass spectrometry)
Chuang 2018[36]	Q1	DBS IDUA activity (tandem mass spectrometry)
Hopkins 2015[25]	Q1	DBS IDUA activity (fluorometric enzymatic assay)
Minter Baerg 2018[35]	Q1	DBS IDUA activity (tandem mass spectrometry)
Aldenhoven 2015a[38]	Q2	HSCT
Aldenhoven 2015b[37]	Q2	HSCT
Eisengart 2018[47]	Q2	ERT
Javed 2018[43]	Q2	HSCT
Kunin-Batson 2016[48]	Q2	HSCT
Langereis 2016[49]	Q2	HSCT
Laraway 2016[44]	Q2	ERT
Megens 2014[45]	Q2	ERT prior to HSCT
Pal 2015[46]	Q2	HSCT or ERT
Poe 2014[42]	Q2	HSCT
Rodgers 2017[40]	Q2	HSCT
Wadhwa 2019[41]	Q2	HSCT
Wyffels 2017[39]	Q2	HSCT/ HSCT + ERT

Abbreviations: DBS: dried blood spot; ERT: enzyme replacement therapy; HSCT: haematopoietic stem cell transplantation; IDUA: α-L-iduronidase.

Publications excluded after review of full-text articles

Of the 110 publications included after the review of titles and abstracts, 93 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 after review of full-text articles

Reference	Reason for Exclusion
Kubaski F, Suzuki Y, Orii K, et al. Glycosaminoglycan levels in dried blood spots of patients with mucopolysaccharidoses and mucolipidoses. Molecular Genetics and Metabolism 2017;120:247-254.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Abdi M, Hakhamaneshi MS, Alaei MR, et al. Validation of urinary glycosaminoglycans in iranian patients with mucopolysaccharidase type i: The effect of urine sedimentation characteristics. Iranian Journal of Child Neurology 2014;8:39-45.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Abdi M, Hakhamaneshi MS, Alaei MR, et al. Determination of Biological Variance and Validation of a Fluorometric Assay for Measurement of alpha-l-Iduronidase Activity in Dried Blood Spots Samples: The First Experience in Iran. Indian Journal of Clinical Biochemistry 2015;30:318-322.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Ahmed A, Rudser K, Kunin-Batson A, et al. Mucopolysaccharidosis (MPS) Physical Symptom Score: Development, Reliability, and Validity. Jimd Reports 2016;26:61-8.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Ahmed A, Shapiro E, Rudser K, et al. Association of somatic burden of disease with age and neuropsychological measures in attenuated mucopolysaccharidosis types I, II and VI. Molecular Genetics and Metabolism Reports 2016;7:27-31.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Aldenhoven M, van den Broek BTA, Wynn RF, et al. Quality of life of Hurler syndrome patients after successful hematopoietic stem cell transplantation. Blood Advances 2017;1:2236-2242.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
An K, Wang Y, Li B, et al. Prognostic factors and outcome of patients undergoing hematopoietic stem cell transplantation who are admitted to pediatric intensive care unit. BMC Pediatrics 2016;16 (1) (no pagination).	Not in a relevant population
Anonymous. Erratum: Aldenhoven M, van den Broek BTA, Wynn RF, et al. Quality of life of Hurler syndrome patients after successful hematopoietic stem cell transplantation. Blood Adv. 2017;1(24):2236-2242. Blood Advances 2017;1:2535.	Not a relevant study type
Aranda CS, Ensina LF, Nunes IC, et al. Diagnosis and management of infusion-related hypersensitivity reactions to enzyme replacement therapy for lysosomal diseases: The role of desensitization. Journal of Allergy and Clinical Immunology: In Practice 2016;4:354-356.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Auray-Blais C, Lavoie P, Tomatsu S, et al. UPLC-MS/MS detection of disaccharides derived from glycosaminoglycans as biomarkers of mucopolysaccharidoses. Analytica Chimica Acta 2016;936:139-148.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Bitencourt FHD, Vieira TA, Steiner CE, et al. Medical Costs Related to Enzyme Replacement Therapy for Mucopolysaccharidosis Types I, II, and VI in Brazil: A Multicenter Study. Value in Health Regional Issues 2015;8:99-106.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Bolourchi M, Renella P, Wang RY. Aortic Root Dilatation in Mucopolysaccharidosis I-VII. International Journal of Molecular Sciences 2016;17:29.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Braunlin E, Miettunen K, Lund T, et al. Hematopoietic cell transplantation for severe MPS I in the first six months of life: The heart of the matter. Molecular Genetics and Metabolism 2019;126:117-120.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Braunlin E, Steinberger J, DeFor T, et al. Metabolic Syndrome and Cardiovascular Risk Factors after Hematopoietic Cell Transplantation in Severe Mucopolysaccharidosis Type I (Hurler Syndrome). Biology of Blood and Marrow Transplantation 2018;24:1289-1293.	Does not report a relevant treatment type or appropriate treatment/screening outcomes

Reference	Reason for Exclusion
Breier AC, Ce J, Coelho JC. Use of a commercial agarose gel for analysis of urinary glycosaminoglycans in mucopolysaccharidoses. Brazilian Journal of Pharmaceutical Sciences 2016;52:693-698.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Burton BK, Charrow J, Hoganson GE, et al. Newborn Screening for Lysosomal Storage Disorders in Illinois: The Initial 15-Month Experience. Journal of Pediatrics 2017;190:130-135.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Camargo Neto E, Schulte J, Pereira J, et al. Neonatal screening for four lysosomal storage diseases with a digital microfluidics platform: Initial results in Brazil. Genetics and Molecular Biology 2018;41:414-416.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Castilhos CD, Mezzalira J, Goldim MPS, et al. Determination of the lysosomal hydrolase activity in blood collected on filter paper, an alternative to screen high risk populations. Gene 2014;536:344-347.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Chan MJ, Liao HC, Gelb MH, et al. Taiwan National Newborn Screening Program by Tandem Mass Spectrometry for Mucopolysaccharidoses Types I, II, and VI. Journal of Pediatrics 2019;205:176-182.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Chennamaneni NK, Kumar AB, Barcenas M, et al. Improved reagents for newborn screening of mucopolysaccharidosis types I, II, and VI by tandem mass spectrometry. Analytical Chemistry 2014;86:4508-14.	Not in a relevant population
Cobos PN, Steglich C, Santer R, et al. Dried blood spots allow targeted screening to diagnose mucopolysaccharidosis and mucolipidosis. Jimd Reports 2015;15:123-32.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Coletti HY, Aldenhoven M, Yelin K, et al. Long-term functional outcomes of children with hurler syndrome treated with unrelated umbilical cord blood transplantation. Jimd Reports 2015;20:77-86.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Colon C, Alvarez JV, Castano C, et al. A selective screening program for the early detection of mucopolysaccharidosis: Results of the FIND project-A 2-year follow-up study. Medicine (United States) 2017;96 (19) (no pagination).	SLR/Not in a relevant population
Conner T, Cook F, Fernandez V, et al. An online survey on burden of illness among families with post-stem cell transplant mucopolysaccharidosis type i children in the United States 11 Medical and Health Sciences 1117 Public Health and Health Services. Orphanet Journal of Rare Diseases 2019;14 (1) (no pagination).	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Darba J, Ascanio M. Enzymatic replacement therapy for lysosomal storage disorders: Drug evaluations review in Spain. Health Policy and Technology 2019;8:14-23.	Not a relevant study type
Dave MB, Chawla PK, Dherai AJ, et al. Urinary Glycosaminoglycan Estimation as a Routine Clinical Service. Indian Journal of Clinical Biochemistry 2015;30:293-297.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Deambrosis D, Lum SH, Hum RM, et al. Immune cytopenia post-cord transplant in Hurler syndrome is a forme fruste of graft rejection. Blood Advances 2019;3:570-574.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Dickson PI, Kaitila I, Harmatz P, et al. Safety of laronidase delivered into the spinal canal for treatment of cervical stenosis in mucopolysaccharidosis I. Molecular Genetics and Metabolism 2015;116:69-74.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Dickson PI, Kaitila I, Harmatz P, et al. Data from subjects receiving intrathecal laronidase for cervical spinal stenosis due to mucopolysaccharidosis type I. Data in Brief 2015;5:71-6.	Not a relevant study type
Dornelles AD, Artigalas O, Da Silva AA, et al. Efficacy and safety of intravenous laronidase for mucopolysaccharidosis type I: A systematic review and meta-analysis. PLoS ONE 2017;12 (8) (no pagination).	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Dornelles AD, De Camargo Pinto LL, De Paula AC, et al. Enzyme replacement therapy for mucopolysaccharidosis type I among patients followed within the MPS Brazil network. Genetics and Molecular Biology 2014;37:23-29.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Dualibi APFF, Martins AM, Moreira GA, et al. The impact of laronidase treatment in otolaryngological manifestations of patients with mucopolysaccharidosis. Brazilian Journal of Otorhinolaryngology 2016;82:522-528.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Eisengart JB, Pierpont EI, Kaizer AM, et al. Intrathecal enzyme replacement for Hurler syndrome: biomarker association with neurocognitive outcomes. Genetics in Medicine 2019;25:25.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Elliott S, Buroker N, Cournoyer JJ, et al. Pilot study of newborn screening for six lysosomal storage diseases using Tandem Mass Spectrometry. Molecular Genetics and Metabolism 2016;118:304-309.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Euctr DE. Treatment of patients, who have Mucopolysaccharidosis Type I, receiving pentosan polysulfate subcutaneous injections weekly. Http://www.who.int/trialsearch/trial2.aspx? Trialid=euctr2014-000350-11-de 2014.	Does not report a relevant treatment type or appropriate treatment/screening outcomes

Reference	Reason for Exclusion
Eyskens F, Devos S. Newborn Screening for Lysosomal Storage Disorders in Belgium: The Importance of Sex- and Age-Dependent Reference Ranges. Journal of Inborn Errors of Metabolism and Screening 2017;5.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Ghosh A, Miller W, Orchard PJ, et al. Enzyme replacement therapy prior to haematopoietic stem cell transplantation in Mucopolysaccharidosis Type I: 10 year combined experience of 2 centres. Molecular Genetics and Metabolism 2016;117:373-377.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Giugliani R, Giugliani L, De Oliveira Poswar F, et al. Neurocognitive and somatic stabilization in pediatric patients with severe Mucopolysaccharidosis Type i after 52 weeks of intravenous brain-penetrating insulin receptor antibody-iduronidase fusion protein (valanafusp alpha): An open label phase 1-2 trial. Orphanet Journal of Rare Diseases 2018;13 (1) (no pagination).	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Giugliani R, Vieira TA, Carvalho CG, et al. Immune tolerance induction for laronidase treatment in mucopolysaccharidosis I. Molecular Genetics and Metabolism Reports 2017;10:61-66.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Gucciardi A, Legnini E, Di Gangi IM, et al. A column-switching HPLC-MS/MS method for mucopolysaccharidosis type I analysis in a multiplex assay for the simultaneous newborn screening of six lysosomal storage disorders. Biomedical Chromatography 2014;28:1131-1139.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Guilheiro JM, Chaves MD, Martins AM, et al. Cytogenetic biomonitoring in mucopolyssacharosis I, II and IV patients treated with enzyme replacement therapy. Toxicology Mechanisms and Methods 2014;24:603-607.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Henley WE, Anderson LJ, Wyatt KM, et al. The NCS-LSD cohort study: a description of the methods and analyses used to assess the long-term effectiveness of enzyme replacement therapy and substrate reduction therapy in patients with lysosomal storage disorders. Journal of Inherited Metabolic Disease 2014;37:939-944.	Not in a relevant population
Hetmanczyk K, Bednarska-Makaruk M, Kierus K, et al. Monitoring of dipeptidyl peptidase-IV (DPP-IV) activity in patients with mucopolysaccharidoses types I and II on enzyme replacement therapy - Results of a pilot study. Clinical Biochemistry. 2015;10.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Hinderer C, Katz N, Louboutin JP, et al. Abnormal polyamine metabolism is unique to the neuropathic forms of MPS: Potential for biomarker development and insight into pathogenesis. Human Molecular Genetics 2017;26:3837-3849.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Hong X, Kumar AB, Ronald Scott C, et al. Multiplex tandem mass spectrometry assay for newborn screening of X-linked adrenoleukodystrophy, biotinidase deficiency, and galactosemia with flexibility to assay other enzyme assays and biomarkers. Molecular Genetics and Metabolism 2018;124:101-108.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Jameson E, Jones S, Remmington T. Enzyme replacement therapy with laronidase (Aldurazyme) for treating mucopolysaccharidosis type I. Cochrane Database of Systematic Reviews 2016;2016 (4) (no pagination).	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Kubaski F, Mason RW, Nakatomi A, et al. Newborn screening for mucopolysaccharidoses: a pilot study of measurement of glycosaminoglycans by tandem mass spectrometry. Journal of Inherited Metabolic Disease 2017;40:151-158.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Kadali S, Patlolla RD, Kolusu A, et al. The utility of two dimensional electrophoresis in diagnosis of mucopolysaccharidosis disorders. Clinica Chimica Acta 2016;457:36-40.	Not in a relevant population
Kato S, Yabe H, Takakura H, et al. Hematopoietic stem cell transplantation for inborn errors of metabolism: A report from the Research Committee on Transplantation for Inborn Errors of Metabolism of the Japanese Ministry of Health, Labour and Welfare and the Working Group of the Japan Society for Hematopoietic Cell Transplantation. Pediatric Transplantation 2016;20:203-214.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Keilmann A, Bendel F, Nospes S, et al. Alterations of mucosa of the larynx and hypopharynx in patients with mucopolysaccharidoses. Journal of Laryngology and Otology 2016;130:194-200.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Koehne T, Kohn A, Friedrich RE, et al. Differences in maxillomandibular morphology among patients with mucopolysaccharidoses I, II, III, IV and VI: a retrospective MRI study. Clinical oral investigations 2018;22:1541-1549.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Kuiper G, Nijmeijer SCM, Roelofs MJM, et al. Limited Data to Evaluate Real-World Effectiveness of Enzyme Replacement Therapy for Mucopolysaccharidosis type I. Journal of Inherited Metabolic Disease 2019;25:25.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Kuiper GA, van Hasselt PM, Boelens JJ, et al. Incomplete biomarker response in mucopolysaccharidosis type I after successful hematopoietic cell transplantation. Molecular Genetics and Metabolism 2017;122:86-91.	Does not report a relevant treatment type or appropriate treatment/screening outcomes

Reference	Reason for Exclusion
Langereis EJ, van Vlies N, Church HJ, et al. Biomarker responses correlate with antibody status in mucopolysaccharidosis type I patients on long-term enzyme replacement therapy. Molecular Genetics and Metabolism 2015;114:129-137.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Langereis EJ, Wagemans T, Kulik W, et al. A multiplex assay for the diagnosis of mucopolysaccharidoses and mucolipidoses. PLoS ONE 2015;10 (9) (no pagination).	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Liao HC, Chiang CC, Niu DM, et al. Detecting multiple lysosomal storage diseases by tandem mass spectrometry - A national newborn screening program in Taiwan. Clinica Chimica Acta 2014;431:80-86.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Lin HY, Chuang CK, Chen MR, et al. Cardiac structure and function and effects of enzyme replacement therapy in patients with mucopolysaccharidoses I, II, IVA and VI. Molecular Genetics and Metabolism 2016;117:431-437.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Lin HY, Lee CL, Lo YT, et al. The relationships between urinary glycosaminoglycan levels and phenotypes of mucopolysaccharidoses. Molecular genetics & genomic medicine 2018;6:982-992.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Lin HY, Shih SC, Chuang CK, et al. Assessment of hearing loss by pure-tone audiometry in patients with mucopolysaccharidoses. Molecular Genetics and Metabolism 2014;111:533-538.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Liu Y, Yi F, Kumar AB, et al. Multiplex tandem mass spectrometry enzymatic activity assay for newborn screening of the mucopolysaccharidoses and type 2 neuronal ceroid lipofuscinosis. Clinical Chemistry 2017;63:1118-1126.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Lum SH, Miller WP, Jones S, et al. Changes in the incidence, patterns and outcomes of graft failure following hematopoietic stem cell transplantation for Hurler syndrome. Bone Marrow Transplantation 2017;52:846-853.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Lum SH, Stepien KM, Ghosh A, et al. Long term survival and cardiopulmonary outcome in children with Hurler syndrome after haematopoietic stem cell transplantation. Journal of Inherited Metabolic Disease 2017;40:455-460.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Maccari F, Galeotti F, Mantovani V, et al. Composition and structure of glycosaminoglycans in DBS from 2-3-day-old newborns for the diagnosis of mucopolysaccharidosis. Analytical Biochemistry 2018;557:34-41.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Maccari F, Galeotti F, Zampini L, et al. Total and single species of uronic acid-bearing glycosaminoglycans in urine of newborns of 2-3 days of age for early diagnosis application. Clinica Chimica Acta 2016;463:67-72.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Madoff LU, Kordun A, Cravero JP. Airway Management in Patients with Mucopolysaccharidoses: The Progression Towards Difficult Intubation. Paediatric anaesthesia. 2019;31.	Not in a relevant population
Makino E, Klodnitsky H, Leonard J, et al. Publisher Correction: Fast, sensitive method for trisaccharide biomarker detection in mucopolysaccharidosis type 1. Scientific Reports 2018;8:4994.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Matsubara Y, Miyazaki O, Kosuga M, et al. Cerebral magnetic resonance findings during enzyme replacement therapy in mucopolysaccharidosis. Pediatric Radiology 2017;47:1659-1669.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Monachesi C, Zampini L, Padella L, et al. False positive screen test for mucopolysaccharidoses in healthy female newborns. Clinica Chimica Acta 2018;486:221-223.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Moreau J, Brassier A, Amaddeo A, et al. Obstructive sleep apnea syndrome after hematopoietic stem cell transplantation in children with mucopolysaccharidosis type I. Molecular Genetics and Metabolism 2015;116:275-280.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Navarrete-Martinez JI, Limon-Rojas AE, Gaytan-Garcia MJ, et al. Newborn screening for six lysosomal storage disorders in a cohort of Mexican patients: Three-year findings from a screening program in a closed Mexican health system. Molecular genetics and metabolism 2017;121:16-21.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Noh H, Lee JI. Current and potential therapeutic strategies for mucopolysaccharidoses. Journal of Clinical Pharmacy and Therapeutics 2014;39:215-224.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Ohira M, Okuyama T, Mashima R. Quantification of 11 enzyme activities of lysosomal storage disorders using liquid chromatography-tandem mass spectrometry. Molecular Genetics and Metabolism Reports 2018;17:9-15.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Osipova LA, Kuzenkova LM, Namazova-Baranova LS, et al. Efficacy and safety of enzyme replacement therapy in children with mucopolysaccharidosis type I, II, and VI: A single-center cohort study. [Russian]. Voprosy Sovremennoi Pediatrii - Current Pediatrics 2018;17:76-84.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Pal AR, Brown N, Jones SA, et al. Obstructive sleep Apnea in MPS: A systematic review of pretreatment and posttreatment prevalence and severity. Journal of Inborn Errors of Metabolism and Screening 2015;2015:1-10.	Does not report a relevant treatment type or appropriate treatment/screening outcomes

Reference	Reason for Exclusion
Pal AR, Mercer J, Jones SA, et al. Substrate accumulation and extracellular matrix remodelling promote persistent upper airway disease in mucopolysaccharidosis patients on enzyme replacement therapy. PloS one 2018;13:e0203216.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Pardridge WM, Boado RJ, Giugliani R, et al. Plasma Pharmacokinetics of Valanafusp Alpha, a Human Insulin Receptor Antibody-Iduronidase Fusion Protein, in Patients with Mucopolysaccharidosis Type I. BioDrugs 2018;32:169-176.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Perez-Lopez J, Morales-Conejo M, Lopez-Rodriguez M, et al. Efficacy of laronidase therapy in patients with mucopolysaccharidosis type I who initiated enzyme replacement therapy in adult age. A systematic review and meta-analysis. Molecular Genetics and Metabolism 2017;121:138-149.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Polgreen LE, Vehe RK, Rudser K, et al. Elevated TNF-alpha is associated with pain and physical disability in mucopolysaccharidosis types I, II, and VI. Molecular Genetics and Metabolism 2016;117:427-430.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Poswar FO, de Souza CFM, Giugliani R, et al. Aortic root dilatation in patients with mucopolysaccharidoses and the impact of enzyme replacement therapy. Heart and Vessels 2019;34:290-295.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Ribas GS, De Mari JF, Civallero G, et al. Validation of a multiplex tandem mass spectrometry method for the detection of selected lysosomal storage diseases in dried blood spots. Journal of Inborn Errors of Metabolism and Screening 2017;5.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Schmidt M, Breyer S, Lobel U, et al. Musculoskeletal manifestations in mucopolysaccharidosis type i (Hurler syndrome) following hematopoietic stem cell transplantation. Orphanet Journal of Rare Diseases 2016;11 (1) (no pagination).	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Shapiro EG, Nestrasil I, Rudser K, et al. Neurocognition across the spectrum of mucopolysaccharidosis type I: Age, severity, and treatment. Molecular Genetics and Metabolism 2015;116:61-68.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Shimada T, Kelly J, LaMarr WA, et al. Novel heparan sulfate assay by using automated high-throughput mass spectrometry: Application to monitoring and screening for mucopolysaccharidoses. Molecular Genetics and Metabolism 2014;Part 2. 113:92-99.	Not in a relevant population
Skrinjar P, Schwarz M, Lexmuller S, et al. Rapid and Modular Assembly of Click Substrates To Assay Enzyme Activity in the Newborn Screening of Lysosomal Storage Disorders. Acs Central Science 2018;4:1688-1696.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Tanyalcin MT. Urinary glycosaminoglycan electrophoresis with optimized keratan sulfate separation using Peltier system for the screening of mucopolysaccharidoses. Journal of Inborn Errors of Metabolism and Screening 2015;2015:1-5.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Tear Fahnehjelm K, Olsson M, Chen E, et al. Children with mucopolysaccharidosis risk progressive visual dysfunction despite haematopoietic stem cell transplants. Acta Paediatrica, International Journal of Paediatrics 2018;107:1995-2003.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Tortorelli S, Turgeon CT, Gavrilov DK, et al. Simultaneous testing for 6 lysosomal storage disorders and x-adrenoleukodystrophy in dried blood spots by tandem mass spectrometry. Clinical Chemistry 2016;62:1248-1254.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Trim PJ, Hopwood JJ, Snel MF. Butanolysis derivatization: improved sensitivity in LC-MS/MS quantitation of heparan sulfate in urine from mucopolysaccharidosis patients. Analytical chemistry 2015;87:9243-9250.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Tylki-Szymanska A, De Meirleir L, Di Rocco M, et al. Easy-to-use algorithm would provide faster diagnoses for mucopolysaccharidosis type I and enable patients to receive earlier treatment. Acta Paediatrica, International Journal of Paediatrics 2018;107:1402-1408.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Vance M, Llanga T, Bennett W, et al. AAV Gene Therapy for MPS1-associated Corneal Blindness. Scientific reports 2016;6:22131.	Not in a relevant population
Verma J, Thomas DC, Kasper DC, et al. Inherited Metabolic Disorders: Efficacy of Enzyme Assays on Dried Blood Spots for the Diagnosis of Lysosomal Storage Disorders. Jimd Reports 2017;31:15-27.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Wang J, Luan Z, Jiang H, et al. Allogeneic Hematopoietic Stem Cell Transplantation in Thirty-Four Pediatric Cases of Mucopolysaccharidosis-A Ten-Year Report from the China Children Transplant Group. Biology of Blood and Marrow Transplantation 2016;22:2104-2108.	Does not report a relevant treatment type or appropriate treatment/screening outcomes
Wasserstein MP, Caggana M, Bailey SM, et al. The New York pilot newborn screening program for lysosomal storage diseases: Report of the First 65,000 Infants. Genetics in Medicine 2019;21:631-640.	Does not report a relevant treatment type or appropriate treatment/screening outcomes

Appendix 3 — Summary and appraisal of individual studies

Data extraction

Table 15. Studies relevant to criterion 4

Study Reference	Burlina 2018
Study Design	Design Prospective screening study Objective To evaluate a multiplexed assay for Fabry disease, Pompe disease, Gaucher disease, and MPS I, and assess its effectiveness for detecting these lysosomal storage disorders (LSDs) in neonates. Dates September 2015 to January 2017 Country Italy Setting Regional North East Italy ENBS program
Population Characteristics	Participant recruitment/source of samples **Pre-pilot phase:** Residual and de-identified dried blood spot (DBS) samples from over 3,500 newborn blood spot screening (NBS) specimens (healthy) were used for the pre-pilot enzymatic activity cut-off determinations. Forty-one patients with confirmed LSDs, including 14 Gaucher, 6 neonatal-onset Pompe, 19 Fabry disease, and 2 MPS I patients were also analysed. The spots were collected after diagnosis and all samples had low enzyme activities that could be clearly distinguished from those of healthy newborns, with the exception of female patients tested for Fabry disease, for whom the enzyme activity assay does not reliably discern heterozygosity. **Screening phase**: DBS samples from newborns were collected consecutively by the Regional North East Italy ENBS program after informed consent was obtained from a parent. Samples were collected on the same card that other NBS tests were collected; a second sample was required for premature babies (<34 gestational weeks and/or weight <2000 g) and for sick newborns (those receiving transfusions or parenteral nutrition). DBS were analysed the day they were received, and then the DBS cards were stored in plastic bags at −5 degrees Celsius for at least 5 years after analysis. Time of sample collection (days after birth) Samples were collected between 36 and 48 hours of life. Prevalence of MPS I in the study One newborn was diagnosed with MPS I (incidence 1/44,411). The incidence of LSDs in total was 1 in 4441 births.

Study Reference | **Burlina 2018**

Sample size

N screened/invited = 44,411
N eligible = 44,411
N screened = 44,411
N excluded (with reason) = 0
N lost to follow-up = 0
N completed = 44,411
N excluded from analysis = 0
N included in analysis = 44,411

Demographics

NR

Screening Method

Index test
Tandem mass spectrometry using the NeoLSD assay system

- The NeoLSD kit contained the buffer, mobile phase, substrates, and internal standards for assaying the DBS activities of 6 enzymes: acid α-glucosidase (GAA; Pompe), acid α-galactosidase (GLA; Fabry), **α-L-iduronidase** (**IDUA; MPS I)**, acid β-glucocerebrosidase (ABG; Gaucher) and acid sphingomyelinase (ASM), and β-galactosidase (GALC).
- Provisional cut-offs were set for the pre-pilot phase and the first 9 months of screening. According to the enzyme activity, 2 cut-off levels were set: a slightly elevated value between 0.2 multiple of the median (MoM) (and the 0.25 percentile (conservative values) and lower values below 0.2 MoM (high-risk value). Both cut-offs were validated with known clinical case controls; the high-risk cut-off value was chosen because it enabled detection of positive cases while avoiding too many recalls from the newborn screening process.
- If the enzymatic value from the index test was below the high-risk cut-off, a second DBS was requested. If the activity of the second spot was still below the cut-off, the infant was referred for confirmatory testing and clinical follow-up.

Reference standard
Different assessments were performed depending on the LSD, including clinical evaluation, mutational analysis (patients and parents), substrate quantification and/or enzyme activities in leukocytes/lymphocytes. For newborns who screened positive for MPS I, urinary GAG analyses, IDUA testing and mutational analyses were performed to confirm diagnoses.

Test Accuracy

Newborn blood spot screening results for MPS I:

Test outcome	Value
Patients with positive initial newborn screening[a]	13
Patients to undergo confirmatory testing	8
Patients with confirmed disorder	1
Pseudodeficiency	5
Variant of unknown significance	0
Carrier (carrier/wild type; carrier/pseudodeficiency)	2

Study Reference	Burlina 2018
	Prevalence: 44,411 **PPV, % (true positive/recalls %): 7.7** [a] Cut-off <0.2 MoM
Authors' Conclusions	In conclusion, the NeoLSD® MS/MS assay system for DBS screening proved effective in identifying neonates at risk for LSDs in this population-based NBS program. Establishing cut-off values before starting a screening program is essential to avoid a high number of false positives, which are a source of needless anxiety and unnecessary medical interventions. Long-term follow-up of the affected infants may provide important information about the natural history of the disease and their specific mutations, and should allow for optimized treatment outcomes.

Abbreviations: ABG: acid β-glucocerebrosidase; ASM: acid sphingomyelinase; DBS: dried blood spot; ENBS: expanded newborn screening; GAA: α-glucosidase; GAG: glycosaminoglycan; GALC: β-galactosidase; GLA: acid α-galactosidase; IDUA: α-L-iduronidase; LSD: lysosomal storage disorder; MoM: multiple of the median; MPS I: mucopolysaccharidosis type I; MS/MS: multiplexed tandem mass spectrometry; NBS: newborn blood spot screening; NR: not reported; PPV: positive predictive value.

Study Reference	Chuang 2018
Study Design	Design Prospective screening study Objective To investigate the current status of several genotypes that may cause pseudo deficiencies in IDS enzyme activity, and to report the positive findings of MPS I and MPS II through confirmatory diagnostic experiments Dates August 2015 to November 2017 Country Taiwan Setting Screening at three Newborn Screening Centers in Taiwan; suspected cases referred to Mackay Memorial Hospital for confirmation
Population Characteristics	Participant recruitment/source of samples Suspected cases in infants with a reduction in either IDUA or IDS enzyme activity in DBS detected by tandem mass spectrometry assay were referred to MacKary Memorial Hospital for MPS confirmation after the first and repeat NBS tests. The samples required for the assay included urine (10–20 mL) and EDTA blood (2 tubes, 3–5 mL in each) Time of sample collection and analysis (days after birth) Sample collection: NR Newborn screening test (DBS analysis): Reported for 4/8 infants who underwent MPS I screening Subject | Age of the test (months)

Study Reference	Chuang 2018

3	1.1
5	1.4
6	1.4
7	4.5

Prevalence of MPS I in the study

4 newborns were diagnosed with MPS I, with a prevalence rate of 1.35 per 100,000 live births

Sample size

N screened/invited = 294,196 (newborn screening [received index test])
N eligible = 8 (suspected cases referred for confirmation)
N screened = 8 (analysed for confirmatory diagnosis)
N excluded (with reason) = 0
N lost to follow-up = 0
N completed = 8
N excluded from analysis = 0
N included in analysis = 8

Demographics

Newborns: n=8

- Female: n=5
- Male: n=3

Screening Method

Index test
IDUA activity tested by tandem mass spectrometry assay. If an abnormal test result was obtained, the NBS test was repeated on the same initial sample. If positive, a second DBS sample was collected and re-tested. The cut-off value was 3.0 μmol/L/h for both the initial and second DBS tests. If the result of the second DBS sample for tandem mass spectrometry was positive, the case was considered to be high-risk and referred for confirmatory testing.

Reference standard
Suspected cases were referred to Mackay Memorial Hospital for confirmatory analysis. Urinary quantitative GAG analysis was confirmed using the dimethylmethylene blue method (DMB/creatinine ratio), two-dimensional electrophoresis and liquid chromatography/tandem mass spectrometry to detect predominant GAG-derived disaccharides (DS+HS, DS only or HS only).
Differential diagnosis was achieved using a leukocyte enzyme assay for β-iduronidase and molecular DNA analysis using PCR to confirm deficiency of IDUA activity and genotype, respectively.

Test Accuracy

Screening results for MPS I:
The estimated recall rate for MPS I was 0.005% (initially tested positive)

Test outcome	Value
Positive predictive value, %	26.7

Study Reference	Chuang 2018
	Patients failed first test (positive test result), n: 15 Patients failed second test (positive test result), n: 8 Patients confirmed diagnosis (true positive), n: 4
Authors' Conclusions	The entire diagnostic process including a first tier newborn screening test by tandem mass spectrometry assay and second tier confirmatory analysis are comprehensive and facile. However, the positive predictive values of MPS I NBS are low.

Abbreviations: DBS: dried blood spot; DMB: dimethylmethylene blue method; DNA: deoxyribonucleic acid; DS: dermatan sulphate; EDTA: ethylenediaminetetraacetic acid; GAG: glycosaminoglycan; HS: heparan sulphate; IDUA: α-iduronidase; IDS: iduronate-2-sulfatase; MPS: mucopolysaccharidoses; NBS: newborn screening; NR: not reported; PCR: polymerase chain reaction.

Study Reference	Hopkins 2015
Study Design	Design Prospective screening study Objective To evaluate the performance of a state-wide full-population pilot study in Missouri on newborn blood spots for screening of lysosomal storage disorders (LSDs) using digital microfluidics Dates January 2013–NR Country United States Setting Missouri State Public Health Laboratory
Population Characteristics	Participant recruitment/source of samples **Pre-pilot phase**: Residual and deidentified dried blood spot samples from over 13,000 specimens from newborns who were routinely screened in Missouri and were stored in accordance with local policy, were used to determine the enzymatic activity cut-offs, along with samples from 29 known clinical cases provided by contracted genetic referral centers. **Screening phase**: Fresh DBS samples were punched from routine newborn blood spot screening (NBS) specimens received at the study institute during the study period. All newborn specimens (n=43,701) which were received during the study in Missouri were screened. All specimens which were collected from premature infants, sick infants, or infants aged <24 hours automatically mandated a repeat screen to avoid false-positive or false-negative results, and unreliable NBS results. Time of screening (days after birth) NR Prevalence of MPS I in the study

Study Reference	Hopkins 2015
	Three newborns were diagnosed with MPS I, with an incidence rate of 1:14,567. Sample size N screened/invited = NA N eligible = 43,701 N screened = 43,701 N excluded (with reason) = NR N lost to follow-up = 3 N completed = 43,698 N excluded from analysis = 4 (confirmatory results were pending for 4 screen-positive MPS I samples) N included in analysis = 43,690 Demographics NR
Screening Method	Index test A multiplexed fluorometric enzymatic assay of newborn dried blood spots on a digital microfluidic platform, to detect four lysosomal storage disorders, including MPS I. The initial cut-off value for MPS I detection was determined using normal patient percentile data obtained from the deidentified pre-pilot samples, along with dried blood spot enzyme-level data from 29 known clinical cases provided by contracted genetic referral centers. The pilot cut-off for referral of positive screens for MPS I was set at 4.0 µmol/L/hour (0.8 percentile) for IDUA. This was chosen to ensure that the known clinical case control samples could be detected without provoking too many referred cases from the newborn screening process, considering the expected incidence of MPS I. Reference standard The diagnosis laboratory tests that were used to confirm diagnoses were not reported. Samples that tested positive (with average triplicate screening values breaching the cut-off) were assessed for risk by reviewing other LSD enzyme results from the multiplex assay along with the infant's gestational age, age at specimen collection, and health status – all newborn blood spot specimens collected from premature infants, sick infants, or infants aged <24 hours automatically mandated a repeat screen, because these circumstances can produce false-positive, false-negative, and unreliable results. If none of these conditions applied and if the quality of the specimen was adequate, the screen-positive result was considered high-risk and referred to genetic referral centres for evaluation, confirmatory testing and diagnosis.
Test Accuracy	**Newborn blood spot screening results for MPS I:**

Test outcome	Value
Positive predictive value, %	11
6-month false positive rate, %	0.037
Patients screened positive, n	32
Patients with confirmed disorder, n	1
Patients with condition of unknown significance or onset, n	2
Patients with pseudodeficiency, n	7

Study Reference	Hopkins 2015	
	Carriers, n	2
	Patients with false-positive results, n	16
	Patients lost to follow-up, n	0
	Patients with status pending, n	4
Authors' Conclusions	The first 6 months of the Missouri LSD pilot study provided the opportunity to validate the effectiveness of the digital microfluidic screening method, refine the cut-offs for detection of these LSDs, and test the entire system of infant referral, follow-up, confirmation, treatment, and screening program communication	

Abbreviations: DBS: dry blood spot; IDUA: alpha-L-iduronidase; LSD: lysosomal storage disorder; MPS I: mucopolysaccharidosis type I; NA: not applicable; NBS: newborn blood spot screening; NR: not reported.

Study Reference	Minter Baerg 2018
Study Design	Design Prospective screening study Objective To report on an informatics solution to minimise issues associated with newborn screening for lysosomal disorders such as poor specificity, psychosocial harm experienced by caregivers, and costly follow-up testing of false-positive cases. Dates February 2016 to February 2017 Country US Setting Kentucky Department for Public Health
Population Characteristics	Participant recruitment/source of samples Newborn blood spot specimens were collected from infants born over one year. Time of screening (days after birth) Not reported for the entire sample. For the single MPS I case, the sample was collected 58 hours after birth. Prevalence of MPS I in the study One patient was diagnosed with MPS I, and a further patient was found to be a carrier. Sample size N screened/invited = 55,161 N eligible = 55,161 N screened = 55,161 N excluded (with reason) = 0

Study Reference	**Minter Baerg 2018**
	N lost to follow-up = NR N completed = 55,161 N excluded from analysis = 0 N included in analysis = 55,161 Demographics NR
Screening Method	Index test **1. First-tier 6-plex assay by tandem mass spectrometry integrated with multivariate pattern recognition software** Six enzyme activities were measured simultaneously by flow-infusion tandem mass spectrometry. The Collaborative Laboratory Integrated Reports (CLIR) single-condition (applicable to diagnosis of one condition) and dual scatter plot tools (applicable to differential diagnosis between 2 conditions with overlapping phenotypes) were used to assess likelihood of disease in 2 stages. First, samples were tested using the 6-plex assay and analysed using the single-condition tool – to minimise the risk of overlooking affected cases, the 6-plex primary screening was set to resolve as negative only cases with a completely normal profile (i.e. all potentially informative markers were within the 1–99% percentile of the respective reference range). The dual scatter plot tool was used to segregate potential true-positive and false-positive cases. Samples which tested 'positive' using both CLIR tools underwent repeat analysis. **2. Repeat analysis with cumulative 10-plex assay of lysophosphatidylcholines and second-tier tests (dermatan sulphate and heparan sulphate concentrations)** When a repeat analysis was indicated to investigate an initial abnormal result, the concentrations of C20–C6 lysophosphatidylcholines were also measured. A second-tier test for the evaluation of dermatan sulphate and heparan sulphate concentrations was also performed. CLIR is an online application that maintains an interactive database of laboratory results from multiple sites, originally developed to improve performance of newborn screening by tandem mass spectrometry. CLIR's defining characteristics are (i) the replacement of analyte cut-off values with condition-specific degree of overlap between cumulative reference and disease ranges, and (ii) the integration of primary markers with all informative permutations of ratios. Ratios calculated between markers not directly related at the biochemical level are particularly helpful in correcting for preanalytical factors and potential analytical bias. An additional and unique feature of CLIR is the replacement of conventional reference intervals with continuous, covariate-adjusted, moving percentiles. Reference standard Samples which screened positive underwent genotyping to confirm the diagnosis: molecular testing of the GALC common 30-kb deletion and Sanger sequencing of the GALC, **IDUA for MPS I**, and GAA genes were performed using clinically available tests.
Test Accuracy	**Newborn blood spot screening results for MPS I:** **Test outcome** / **Value** Cases identified as positive by the single-condition tool: 76 Cases requiring a repeat analysis and a second-tier test[a]: 57 Cases informative by repeat analysis and/or by second-tier test and therefore reported as screen-positive: 2 Confirmed true positives: 1

Study Reference	Minter Baerg 2018
	Confirmed false positives: 1 **Detection rate**: 1:55,161 **False-positive rate**: 0.0018% **Positive predictive value**: 50% [a] Counts include additional cases where a noninformative resolution by the 6-plex dual scatter plot was overruled for sensitivity verification purposes.
Authors' Conclusions	We report for the first time that automated integration of covariate-adjusted reference intervals and population results combined with second-tier tests can improve the false positive rate of newborn screening for lysosomal disorders to a sustainable, near-zero level. An additional novel element of this work is the curation of condition ranges for false positive cases with heterozygous or pseudo-deficiency genotypes. Postanalytical interpretive tools can drastically reduce false-positive outcomes, with preliminary evidence of no greater risk of false-negative events, still to be verified by long-term surveillance.

Abbreviations: CLIR: The Collaborative Laboratory Integrated Reports; GAA: acid α-glucosidase; GALC: galactocerebrosidase; HSCT: haematopoietic stem cell transplantation; IDUA: α-L-iduronidase; MPS I: mucopolysaccharidosis type I; NR: not reported.

Table 16. Studies relevant to criterion 9

Study Reference	Aldenhoven 2015a
Study Design	Design Retrospective chart review Objective To identify predictors of the long-term outcome of patients with MPS-IH after successful HCT. Dates January 1985 to February 2011 Country Europe and the United States Setting Ten participating centres
Population Characteristics	Patient recruitment/eligibility Patients with MPS IH who received an allogeneic-HCT in one of the 10 participating centres were eligible. Patients with an attenuated phenotype (Hurler-Scheie) were excluded on the basis of the age of diagnosis, genotype, and neurodevelopmental presentation. All patients included in the study had at least 10% donor chimerism. Diagnosis of MPS I Assays of leukocyte IDUA activity at presentation in combination with the clinical phenotype confirmed the diagnosis in all patients.

Duration of follow-up

All included patients had a minimum follow up of 3 years post-HCT.

Sample size
N screened/invited = NR
N eligible = 222
N excluded (with reason) = 5 (attenuated phenotype)

N lost to follow-up = 0
N completed = 217
N excluded from analysis = NR (data obtained from patients who received growth hormone treatment were excluded from the start of treatment)
N included in analysis = NR

Data collection
The medical records of all included patients were retrospectively evaluated. On the basis of medical records as well as the various involved specialists, endpoints were scored according to their presence, and progression was compared with the pre-HCT status and timing of interventions.

- Standardised and validated tests were used to assess neurodevelopmental outcome and produce a developmental quotient/intelligence quotient (DQ/IQ) to demonstrate cognitive impairment levels.
- Growth data collected included weight, height, head circumference and body mass index.
- Neurological endpoints were evaluated using radiologic imaging.
- Orthopaedic endpoints and their surgical intervention were evaluated using radiologic imaging, electrophysiological tests and the involved orthopaedic specialists.
- Cardiac endpoints were based on cardiac ultrasounds and the involved cardiologists.
- Respiratory endpoints were based on polysomnography and the involved paediatricians and ear, nose, and throat specialists.
- Ophthalmologic endpoints and their interventions were measured by eye specialists.
- Audiologic endpoints consisted of the presence of a defined hearing loss and the need for hearing aids based on audiologic tests.
- Endocrinologic endpoints included GH treatment and hypothyroidism requiring treatment.

Data analysis

The association between the various patient, donor, and transplantation-related predictors and the primary endpoints were analysed using linear mixed models. For secondary endpoints, univariate and multivariate regression analysis were used: Cox proportional hazards regression analysis in case of clear event-time endpoints and logistic regression analysis in case of binary endpoints. Univariate predictors of outcome parameters that were statistically significant ($P < 0.1$) were selected for multivariate analysis. Results were expressed as estimate (β), hazard ratios (HRs), or odds ratios (ORs) and corresponding 95% confidence intervals (95% CIs). P values <0.05 were considered statistically significant. Cumulative incidence curves were used to depict event-time endpoints.

Demographic characteristics

Characteristic	Value
Median age at diagnosis, months (range)	9 (0–42)
Age at intervention	

Study Reference	Aldenhoven 2015a

Median age, months (range)	16 (2–47)
<16 months	NR
≥16 months	NR
Median follow-up age, years (range)	9 (3–23)
Gender, n (%)	
Male	122 (56)
Female	NR
Caucasian ethnicity, n (%)	198 (91)
MPS I syndrome, n (%)	
Hurler syndrome	217 (100)
Scheie syndrome	NA
Hurler-Scheie syndrome	NA
Transplantation characteristics, n (%)	
Number of HCT (1/2/3)	179/36/2 (83/16/1)
Donor characteristics, n (%)	
Source (CB/BM/PBSC)	85/118/14 (39/54/7)
Related	73 (34)
Carrier	39 (19)
Previous treatment	
Patients who had received ERT, n (%)	45 (21)
Iduronidase level* <reference, n (%)	55 (26)
Iduronidase level (% of mean), median (range)	82 (13–302)

*Measured in leukocytes

Intervention

Intervention and comparator
Intervention: allogeneic-HCT

Comparator: disease progression was compared with the pre-HCT status, and timing of interventions. For neurodevelopmental outcomes, age equivalents were used to permit comparisons across tests and to identify newly acquired skills. The results were compared with norms for typically developing children.

Outcomes Measured

Primary outcome

Neurodevelopmental outcome: cognitive development where a developmental quotient/intelligence quotient of ≥85 = normal, 70–85 = mild cognitive impairment, 55–70 = moderate cognitive impairment and <55 = severe cognitive impairment.

Growth outcome: weight, height, head circumference and body mass index.

Secondary outcomes

Neurological endpoints: hydrocephalus and cerebral atrophy

Orthopaedic endpoints:

- Thoracolumbar kyphosis
- Cord compression

Study Reference	Aldenhoven 2015a
	• Cervical instability • Hip dysplasia with (sub)luxation • Genu valgum • Carpal tunnel syndrome • Trigger fingers **Respiratory endpoints:** Overnight hypoxia and need for respiratory support **Ophthalmologic endpoints:** • Corneal clouding • Glaucoma • Cataracts **Audiologic endpoints:** Presence of a defined hearing loss and need for hearing aids **Endocrinologic endpoints:** GH treatment and hypothyroidism requiring treatment
Effectiveness of the Intervention	Efficacy Primary outcomes: **Neurodevelopmental outcome:** Higher age at HCT (β: -8.40; 95% CI: -14.62 to -2.19; P=.009) was a statistically significant predictor of inferior neurodevelopmental outcome post-HCT. Secondary outcomes: **Neurological endpoints:** Multivariate logistic regression analyses showed age at HCT determined cerebral atrophy after HCT (median age at HCT <16 months, 23%; OR: 1; median age at HCT ≥16 months, 46%; OR: 3.22; 95% CI: 1.60 to 6.50; P=.001). **Orthopaedic endpoints:** Multivariate Cox regression analysis for event-time endpoints showed age at HCT determined cord compression after HCT (median age at HCT <16 months, 5%; HR: 1; median age at HCT ≥16 months, 16%; HR: 2.84; 95% CI: 1.02 to 1.41; P=.04), and carpal tunnel syndrome (median age at HCT <16 months, 33%; HR:1; median age at HCT ≥16 months, 56%; HR:1.72; 95% CI: 1.11 to 2.68; P=.02). **Cardiac endpoints:** Multivariate logistic regression analysis for binary endpoints showed age at HCT determined progression of mitral valve insufficiency (median age at HCT <16 months, 26%; OR: 1; median age at HCT ≥16 months, 47%; OR: 2.46; 95% CI: 1.30 to 4.65; P=.006) and aortic valve insufficiency (median age at HCT <16 months, 19%; OR: 1; median age at HCT ≥16 months, 37%; OR: 2.40; 95% CI: 1.19 to 4.82; P=.01). **Related subgroup analysis**: Combining the predictors age at HCT and baseline DQ/IQ shows that 71.1% of the patients with an age at HCT younger than 12 months in combination with a baseline DQ/IQ lower than 70 develop severe cognitive impairment (DQ/IQ<70) compared with 14.7% if the age at HCT is younger than 12 months combined with a baseline DQ/IQ higher than 70 (P<0.001) Safety NR
Authors' Conclusions	Age at HCT, obtained IDUA level post-HCT and baseline clinical status were all important predictors for the prognosis of patients with MPS-IH post-HCT. Identification of the importance of age at HCT and delivered IDUA enzyme can therefore lead to improvements in the long-term

Study Reference	Aldenhoven 2015a
	clinical outcomes of transplanted patients with MPS-IH. A higher age at HCT predicted inferior neurodevelopmental outcome, however, a clear cut-off baseline DQ/IQ or age at HCT that predicted moderate or severe cognitive impairment after transplant could not be found.

Abbreviations: BM: bone marrow; CB: cord blood; CI: confidence interval; DQ/IQ: developmental quotient/intelligence quotient; ERT: enzyme replacement therapy; GH: growth hormone; HCT: haematopoietic cell transplantation; HR: hazard ratio; IDUA: α-L-iduronidase; MPS-IH: mucopolysaccharidosis type I-Hurler syndrome; NR: not reported; OR: odds ratio; PBSC: peripheral blood stem cells.

Study Reference	Aldenhoven 2015b
Study Design	Design Prospective cohort study Objective To evaluate the survival and graft outcomes of haematopoietic stem cell transplantation (HSCT) in MPS patients, complying with the international guidelines, in 2 centres performing the highest numbers of HSCTs in mucopolysaccharidosis (MPS) patients in Europe. Dates December 2004 to March 2014 Country The Netherlands and the UK Setting The University Medical Center Utrecht (UMCU) and the Royal Manchester Children's Hospital (RMCH)
Population Characteristics	Patient recruitment/eligibility MPS patients consecutively treated at one of the 2 participating centres, according to the European group for Blood and Marrow Transplantation guidelines for HSCT in MPS patients, were included in the study. Diagnosis of MPS I NR Duration of follow-up Median follow-up was 36 months post-HSCT (range 1 to 93) Sample size N screened/invited = NR N eligible = 62 MPS N excluded (with reason) = 0 N lost to follow-up = 0 N completed = 62 MPS N excluded from analysis = 0 N included in analysis = 62 MPS (56 MPS type I–Hurler)

Study Reference	**Aldenhoven 2015b**

Data collection
Outcomes were measured in patients receiving HSCT at one of the two participating centres.

Data analysis

For predictor analysis, patient (gender, diagnosis, age at HSCT) and HSCT-related (HSCT centre, conditioning regimen, donor type, HLA disparity, total nucleated cells infused) factors were selected. The association between these factors and the primary and secondary endpoints were analysed using Cox proportional hazards regression analysis. Univariate predictors of endpoints with $P < 0.10$ were selected for multivariate analysis. Predictors with $P < 0.05$ in multivariate analysis were considered statistically significant. Kaplan-Meier curves were used to depict outcome probabilities.

Demographic characteristics

Characteristic	N=62
Age at diagnosis	NR
Age at HSCT	
Median age, months (range)	13.5 (3–44)
≤18 months, n	NR
>18 months, n	NR
Gender, n (%)	
Male	37 (59.7)
Female	NR
Race, n (%)	NR
MPS I syndrome, n (%)	
Hurler syndrome	56 (90.3)
Scheie syndrome	0
Hurler-Scheie syndrome	0
Transplant donor type, n (%)	
Unrelated cord blood	41 (66.1)
Unrelated bone marrow or unrelated peripheral blood stem cells	4 (6.5)
Matched sibling donor	17 (27.4)
Previous treatment	
MPS I patients who received peritransplant ERT, n (%)	56 (100)
Iduronidase level	NR

Intervention

Intervention
Conditioning regimens:

- Busulfan + cyclophosphamide followed by Thymoglobulin or alemtuzumab from December 2004 to January 2009
- Fludarabine + busulfan followed by Thymoglobulin or alemtuzumab from January 2009 to March 2014

Study Reference	Aldenhoven 2015b
	Donor hierarchy: Participants underwent unrelated cord blood (UCB) or unrelated/matched sibling bone marrow (BM) or peripheral blood stem cell (PBSC) transplants. A non-carrier matched sibling donor (MSD), matched unrelated cord blood (UCB), or matched unrelated donor (MUD) were considered to be preferred donors. If not available, a mismatched UCB donor was used. Comparator Outcomes were compared between HSCT types (UCB/BM/PBSC) and donor types.
Outcomes Measured	Primary outcomes • Overall survival: defined as survival from HSCT to last contact or death. • Event-free survival: defined as survival from HSCT to last contact, death, autologous reconstitution (<10% donor-derived engraftment), or graft-failure (lack of neutrophil recovery or transient engraftment of donor cells after HSCT and/or requirement for a second HSCT). • The association between age at treatment and the primary outcomes were not reported. Secondary outcomes • Neutrophil engraftment: defined as the first day of achieving neutrophil count >0.5 x 10^9/L for 3 consecutive days. • Platelet engraftment: defined as achieving platelet count >50 x 10^9/L for 7 consecutive days. • **Acute GVHD (aGVHD):** Grades II to IV aGVHD was graded according to published criteria. • **Chronic GVHD (cGVHD):** both limited and extensive cGVHD were graded according to standard criteria and evaluated in patients who survived at least 100 days with sustained engraftment. Veno-occlusive disease and viral reactivation of cytomegalovirus (CMV), adenovirus and Epstein-Barr virus with a viral load >1000 cm/mL were recorded. Urinary glycosaminoglycan excretion below the local upper reference limit was considered normal. • Veno-occlusive disease and viral reactivation of cytomegalovirus, adenovirus, and Epstein-Barr virus • Donor chimerism: a donor chimerism >95% was considered as full donor. • Enzyme level: an enzyme level above the local lower reference limit was considered normal. • Urinary glycosaminoglycan excretion: measurements below the local upper reference limit were considered normal.
Effectiveness of the Intervention	Efficacy A higher age at HSCT was a predictor for aGVHD (HR = 1.13; 95% CI = 1.05–1.21; P = 0.001), cGVHD (HR = 1.08; 95% CI = 1.02–1.15; P = 0.01) and CMV reactivation (HR = 1.09; 95% CI = 1.01–1.18; P = 0.02). Related subgroup analysis (if applicable): NA Safety The secondary endpoints for which age at HSCT was a predictor, were safety-related.
Authors' Conclusions	If complying with the international HCT guidelines, HCT in MPS patients results in high safety and efficacy. This allows extension of HCT to more attenuated MPS types. Because a younger age at HCT is associated with reduction of HCT-related toxicity, newborn screening may further increase safety.

Abbreviations: aGVHD: acute graft versus host disease; BM: bone marrow; cGVHD: chronic graft versus host disease; CI: confidence interval; CMV: cytomegalovirus; EFS: event-free survival; ERT: enzyme replacement therapy; HCT: haematopoietic cell transplantation; HLA: human leukocyte antigen; HR: hazard ratio; MPS: mucopolysaccharidosis; MSD: matched sibling donor; MUD: matched unrelated donor; NA: not applicable; NR: not reported; OS: overall survival; PBSC: peripheral blood stem cell; RMCH; Royal Manchester Children's Hospital; UCB: unrelated cord blood; UK: United Kingdom; UMCU: University Medical Center Utrecht.

Study Reference	**Eisengart 2018**
Study Design	Design Retrospective chart review Objective To examine the long-term outcomes of ERT monotherapy in a unique international cohort of patients with Hurler syndrome who were treated exclusively with ERT from a young age, before significant disease progression was apparent. Dates NR Country International (UK, France, Germany, the Netherlands, Brazil and US) Setting International study
Population Characteristics	Patient recruitment/eligibility Eligibility criteria: MPS-IH patients who were treated exclusively and continuously with ERT from a young age until last follow-up or death. Patients had to have a genotypic diagnosis of MPS-IH, initiated ERT before age 5 years, and no history of HCT. Recruitment: Patients were identified from 3 sources including a larger international study (UK, France, Germany and the Netherlands) Diagnosis of MPS I Clinical diagnosis of MPS-IH was confirmed with genotyping (reported previously for all patients except 2 that were confirmed). Duration of follow-up The study evaluated 10-year follow-up data on a group of patients whose 1-year treatment response to early initiated ERT was previously published (CNS outcomes measured at 14 years). Sample size N screened/invited = NR N eligible = 93 (ERT group:18, HCT group: 54, historical control group: 21) N excluded (with reason) = NR N lost to follow-up = 0 N completed = 93 N excluded from analysis: Survival = 0 (1 patient received treatment after age 5 so was censored and added to the Untreated group for analyses)

Study Reference **Eisengart 2018**

CNS outcomes = 27 (all untreated patients (n=21) were excluded because records on CNS outcomes were unavailable; 1 ERT patient was excluded from hydrocephalus analysis and 5 ERT patients were excluded from cervical cancer cord compression because records on these outcomes were not available)

N included in analysis:

Survival = 93 (63 included in sub-analysis for patients under 3 years)

CNS outcomes = 66

Data collection

The charts of patients were retrospectively reviewed for endpoints relating to the lethality and neurologic pathology of MPS-IH. These data on survival and emergence of hydrocephalus and cervical spinal cord compression were reported from the international centres. Records on CNS outcomes were not available for the untreated historical controls. Both outcomes were defined according to standard clinical guidelines at each institution, including neuroimaging.

Data analysis

Statistics: Descriptive statistics were tabulated per treatment group. Continuous variables were summarized with mean and range while categorical variables were summarized with frequency and percentage. Survival curves were evaluated using Kaplan–Meier estimates while comparisons between groups were based on hazard ratios as estimated by unadjusted Cox proportional hazards models and robust variance estimation. Hydrocephalus and cervical cord compression were evaluated separately based on cumulative incidence functions.

Sensitivity analysis: a sensitivity analysis was conducted in which age at treatment was restricted to younger than 3 years.

Demographic characteristics

Characteristic	Untreated	ERT	HCT
Full sample, n	23	18	54
MPS I syndrome, n (%)			
Hurler syndrome	NR	18 (100)	54 (100)
Age at intervention			
Mean age, years (SD)	NA	2.6 (1.3)	1.5 (0.80)
Median age, years (range)	NA	2.6 (0.5–4.7)	1.3 (0.4–4.8)
Male, n (%)	12 (52)	12 (67)	30 (56)
Under 3 years at initiation, n	NA	10	53
Age at intervention			
Mean age, years (SD)	NA	1.6 (0.74)	1.4 (0.65)
Median age, years (range)	NA	1.3 (0.5–2.7)	1.3 (0.4–2.9)
Male, n (%)	NR	6 (60)	30 (57)
Previous treatment			
HCT patients who had received ERT, n (%)	NA	42 (78)	NA

Intervention

Intervention and comparator

Intervention: continuous ERT from a young age. All patients received the standard dose of ERT except for one previously reported, whose dose was increased.

Study Reference	Eisengart 2018
	Comparators: HCT group – MPS-IH patients transplanted since 2002. Historical control group – historical patients who did not receive any treatment due to lack of available HCT donors in the pre-ERT era.
Outcomes Measured	Primary outcome Survival and CNS outcomes (age at emergence of hydrocephalus and cervical cord compression). Secondary outcomes Outcomes were not split into primary and secondary.
Effectiveness of the Intervention	Efficacy Survival was significantly worse in the Untreated group compared with the ERT group (HR = 2.3; P = 0.008), but better in the HCT group compared with the ERT group (HR = 2.6; P = 0.033). When age at treatment was restricted to <3 these survival differences remained qualitatively similar for the Untreated group compared with the ERT group and for the ERT group compared with the HCT group (HR = 2.4; P = 0.046 and HR = 2.50; P = 0.089, respectively). The P value was slightly higher as expected due to a smaller sample size. The median age of death of the Untreated group was 6.4 years, whilst that of the ERT group was 9.0 years. This did not change significantly if ERT was initiated prior to age 3 (8.9 years). Restricting age at treatment to <3 years did not significantly change CNS outcomes. The cumulative incidence of hydrocephalus in the ERT group was 40% compared with 0% in the HCT group (P = 0.010), while the cumulative incidence of cord compression in the ERT group was 67% compared with 16% in the HCT group (P = 0.013). Related subgroup analysis (if applicable): NA Safety NR
Authors' Conclusions	By addressing the lack of long-term outcome comparison between ERT monotherapy and HCT in Hurler syndrome, this study reveals superior outcomes for survival and CNS pathology with HCT. It also provides clinical data to suggest benefit of ERT and to support existing presumptions that the blood–brain barrier is impermeable to standard doses of ERT.

Abbreviations: CNS: central nervous system; ERT: enzyme replacement therapy; HCT: haematopoietic cell transplantation; HR: hazard ratio; MPS-IH: mucopolysaccharidosis type I-Hurler syndrome; NA: not applicable; NR: not reported; UK: United Kingdom; US: United States; SD: standard deviation.

Study Reference	Javed 2018
Study Design	Design Retrospective cohort study Objective To determine whether the efficacy of HSCT or the age of treatment effected the severity of the ocular phenotype in patients with MPS I Hurler.

Study Reference	Javed 2018
	Dates NR Country UK Setting Manchester Royal Eye Hospital
Population Characteristics	Patient recruitment/eligibility Patients with MPS I Hurler Syndrome who had underwent previous HSCT were identified from the hospital clinic database. Patients were excluded if they had poor ophthalmic follow-up, no recent enzyme data available (last 5 years) and any incomplete data. Diagnosis of MPS I NR Duration of follow-up NR Sample size N screened/invited = 81 N eligible = 35 N excluded (with reason) = 9 (2 had a milder Scheie phenotype, 7 had incomplete data) N lost to follow-up = 0 N completed = 26 N excluded from analysis = 0 N included in analysis = 26 Data collection Data on ocular phenotype were collected retrospectively from the patients' notes at the study institution. Level of corneal clouding (documented by subjective clinical assessment score) and LogMAR visual acuity was recorded for each patient. Corneal clouding was graded in the clinical notes as mild, moderate or severe. Data were collected from the most recent clinic visit. The mean and standard deviation of the iduronidase enzyme level for each level of severity of corneal clouding was calculated. Data analysis Descriptive methods were used to describe the trend and data for the groups under and over 18 months, as the patient numbers were too low for any significant statistical analysis. Demographic characteristics

Characteristic	Value
Age at diagnosis	NR

Study Reference	**Javed 2018**

Age at intervention	
Mean age, months (range)	12.7 (4 to 24)
≤18 months, n	20
>18 months, n	6
Gender, n (%)	
Male	NR
Female	NR
Race, n (%)	NR
MPS I syndrome, n (%)	
Hurler syndrome	26 (100%)
Scheie syndrome	NA
Hurler-Scheie syndrome	NA
Previous treatment	NR
Patients who had received ERT, n	11
Iduronidase level	NR*

*and other baseline characteristics reported for individual cases only.

Intervention

Intervention and comparator

HSCT: Patient data were analysed in 3 groups: patients who received HSCT treatment at or under 18 months old (n=20), those treated over 18 months old (n=6) and all patients combined (n=26).

Outcomes Measured

Primary outcome

- Assessment of visual acuity
- Severity of corneal clouding
- Presence of optic neuropathy or retinopathy
- Biomarker assessment included dermatan sulphate/chondroitin sulphate (DS/CS) ratio and iduronidase level

Secondary outcomes

Primary and secondary outcomes were not specified.

Effectiveness of the Intervention

Efficacy

Corneal clouding and mean iduronidase level for patients treated before and after 18 months of age are presented graphically.

DS/CS ratios, which were significantly higher in patients with severe corneal clouding (p=0.043), were also significantly higher in patients treated under 18 months (p=0.023).

Related subgroup analysis:

Study Reference	Javed 2018
	Iduronidase enzyme levels were significantly lower in the severe corneal clouding group, compared to the mild and moderate group combined (p=0.023). This difference remained even when separating the groups into those treated before or after 18 months (p=0.02). Safety NR
Authors' Conclusions	The finding that visual acuity and corneal clouding are related to enzyme levels and GAG ratios following HSCT is important, as this may have implications for optimising future treatment of patients with MPS. Although numbers are low in the study, this finding would be in keeping with long-term treatment outcomes in other organs following HSCT as reported in the largest Hurler outcome study (Aldenhoven et al. 2015a), demonstrating that patients with the highest enzyme levels post-HSCT have the greatest clinical benefits.

Abbreviations: DS/CS: dermatan sulphate/chondroitin sulphate; GAG: glycosaminoglycan; HSCT: haematopoietic stem cell transplantation; MPS: mucopolysaccharidosis; NA, not applicable; NR: not reported.

Study Reference	Kunin-Batson 2016
Study Design	Design Observational study Objective To examine the cognitive status, adaptive functioning, and quality of life of 47 individuals with MPS-IH between 1 and 24 years status-post HCT, and to evaluate the relative influence of demographic, genetic, and transplant-related factors on these key outcomes. Dates 2004 to 2012 Country United Kingdom and United States Setting Royal Manchester Children's Hospital or Great Ormond Street Hospital (UK), or University of Minnesota (US)
Population Characteristics	Patient recruitment/eligibility UK: participants with MPS-IH who were members of the Society for MPS and Related Diseases in the United Kingdom and previously transplanted at either the Royal Manchester Children's Hospital or Great Ormond Street Hospital were invited to participate in this study. US: Patients diagnosed with MPS-IH who were at least 1 year from having completed HCT and at least partially engrafted were selected from the pool of patients followed in the University of Minnesota clinics after transplant. Participants had to be diagnosed with MPS-IH and at least 1 year from having completed HCT, transplanted between 1985 and 2007, as well as being at least partially engrafted. Incomplete data, missing genetic mutation information and significant health issues were reasons for exclusion. Diagnosis of MPS I

Study Reference	**Kunin-Batson 2016**

Diagnosis of MPS-IH was made in accordance with the clinical guidelines at the time, and all patients included in this study manifested early onset of severe symptoms prior to 2 years of age.

Duration of follow-up

NR

Sample size

N screened/invited = 66 (49 UK and 17 US)
N eligible = 47 (34 UK and 13 US)
N excluded (with reason) = UK – 15 (2 [incomplete data] + 1 [no consent] + 2 [significant health issues] + 7 [less than 1 year from HCT] + 3 [missing genetic mutation information]); US – 4 (lack of genetic mutation information)

N lost to follow-up = 0
N completed = 47
N excluded from analysis = 0
N included in analysis = 47

Data collection

Several measures of cognitive function were used, with the scales differing depending on the age group of the child. Full-scale IQ scores were collapsed across instruments to index cognitive ability with a mean of 100 and a standard deviation of 15 representing the average range of functioning and scores more than 2 standard deviations below the mean (i.e., <70) representing impairment. The Vineland Adaptive Behavior Scales were used to measure children's daily functioning. Domains of the scale are combined to form an adaptive behaviour composite score. Domain areas and the composite score have a mean of 100 and a standard deviation of 15 with 2 standard deviations below the mean representing clinical impairment. The Child Health Questionnaire was used to measure physical and psychosocial well-being. Individual domain scores are aggregated to derive 2 summary component scores: the physical functioning and psychosocial health summary scores which are then converted into norm-referenced T-scores with a mean of 50 and a standard deviation of 10. Poor quality of life has been defined as 2 standard deviations below the mean of the normative sample or a physical functioning or psychosocial health summary score <30. The Hollingshead and Redlich classification of socioeconomic status was used. Genotype was obtained from medical records and classified as known severe and other.

Data analysis

Descriptive characteristics were examined by mutation type and in aggregate. Unadjusted comparisons were based on t-test with unequal variance and Welch degrees of freedom. Separate multiple linear regression models were used to examine the adjusted influence of demographic (sex), treatment factors (age at transplant, time since transplant, radiation treatment, type of transplant, number of transplants and genetic factors) on cognitive, adaptive functioning composite and physical and psychosocial quality of life. All analyses were adjusted for treatment centre and regression analysis of adaptive behaviour functioning was further adjusted for version type. Robust variance estimation was used for confidence intervals and p-values.

Demographic characteristics

Characteristic	Value
Age at diagnosis	NR
Age at time of evaluation	
Mean age, years (SD)	10.5 (6.8)

Study Reference	Kunin-Batson 2016	
	Age at transplant	
	Mean age, months (SD)	18.6 (8.2)
	≤18 months, n	NR
	>18 months, n	NR
	Time since transplant	
	Mean, years (SD)	9.0 (6.7)
	Gender, n (%)	
	Male	29 (61.7)
	Female	18 (38.3)
	Race, n (%)	NR
	MPS I syndrome, n (%)	
	Hurler syndrome	47 (100)
	Scheie syndrome	0
	Hurler-Scheie syndrome	0
	Previous treatment	
	Patients who had received ERT, n	2
	Iduronidase level	NR
	Total body irradiation, n (%)	11 (23.4)
	Transplant type, n (%)	
	HCT-related marrow	26 (55.3)
	HCT-unrelated marrow	14 (29.8)
	Cord blood	7 (14.9)
	Number of transplants, n (%)	
	Single	32 (68.1)
	Two	15 (31.9)
	Engraftment, n (%)	
	≤90%	9 (19.1)
	>90%	33 (70.2)
	Missing	5 (10.6)
Intervention	Intervention and comparator Intervention: HSCT Comparator: Children with severe mutations were compared with those with other mutations. Correlation of transplant-related and demographic variables with the outcomes was analysed.	
Outcomes Measured	Primary outcome • Cognitive functioning: gross motor skills, personal-social development, hand and eye coordination and performance were measured resulting in a general developmental quotient. At the University of Minnesota, full-scale IQ scores were measured • Adaptive behaviour: personal and social sufficiency in the areas of communication, daily living skills, socialisation and motor function were tested	

Study Reference	Kunin-Batson 2016
	• Quality of life including 14 physical and psychosocial domains • Socioeconomic status • Genotype Secondary outcomes Outcomes were not separated into primary and secondary
Effectiveness of the Intervention	Efficacy Age at transplant was not a significant predictor of cognitive (IQ) and adaptive functioning, in the multivariate model. Age at transplant was significantly associated with physical quality of life, with older age at transplant associated with poorer physical functioning (β*: -8.10, 95% CI -13.16 to -3.05, p = 0.002). Related subgroup analysis (if applicable): NA Safety NR
Authors' Conclusions	Mutation type (i.e., homozygous for nonsense or deletion mutations or heterozygous for a combination of these) is significantly associated with both cognitive and functional adaptive outcomes post-transplant and may have relevance for early identification of children at risk for severe long-term neurocognitive impairment despite treatment.

*Measure not specified but assumed to be the β coefficient based on study-reported statistical methodology. Abbreviations: HCT: haematopoietic cell transplant; IQ: intelligence quotient; MPS-IH: mucopolysaccharidosis type I Hurler syndrome; NA: not applicable; NR: not reported; SD, standard deviation; UK: United Kingdom; US: United States.

Study Reference	Langereis 2016
Study Design	Design Retrospective chart review/observational study Objective To describe in detail the course of hip dysplasia in the group of patients, as assessed by radiographic analysis, and to identify potential outcome predictors. Dates NR Country The Netherlands, Ireland, the United Kingdom and the United States Setting Academic Medical Center, Amsterdam, and Wilhelmina Children's Hospital, Utrecht, the Netherlands; Our Lady's Children's Hospital, Dublin, Ireland; Royal Manchester Children's Hospital, Manchester, United Kingdom; and University of Minnesota, Minneapolis, Minnesota.

Study Reference	Langereis 2016
Population Characteristics	Patient recruitment/eligibility Recruitment: • All patients with mucopolysaccharidosis type I Hurler phenotype (MPS I-H) who were known at one of the participating centres and had previously undergone HSCT were considered for inclusion Inclusion criteria: • Diagnosis of MPS I-H • Successful HSCT at an age of no more than 2.5 yr, determined by a post-transplantation donor chimerism of ≥10% • The patient's parent or legal guardian understands the full nature and purpose of the study, and provides informed consent (unless the requirement for informed consent is waived by local regulations) prior to extraction of the patient's data through chart review Exclusion criteria: • A post-HSCT follow-up duration of <2 years • The last successful HSCT received after the age of 2.5 yr or before 1st January 1995 • Availability of <2 radiographic studies, performed prior to hip surgery, in adequate Digital Imaging and Communications in Medicine (DICOM) format Diagnosis of MPS I Determined through documented deficiency of alpha-L-iduronidase (IDUA) enzyme in reference to local laboratory standards and documented Hurler-related mutations in both alleles of the IDUA gene. Duration of follow-up At least greater than 2 years after HSCT. Sample size N screened/invited = 206 N eligible = 52 N excluded (with reason) = 154 (14 because duration of follow-up after HSCT <2 years, 60 because the molecular diagnosis was unavailable, 9 because the genotype was not consistent with the document Hurler-related mutations from the study protocol, seven because of an age of >2.5 years at the time of HSCT, 6 because the HSCT was performed before 1995 and 58 because fewer than two radiographs were available). N lost to follow-up = NR N completed = 52 N excluded from analysis = 0 (only individual data) N included in analysis = 52

Data collection

For the patients who were eligible for inclusion, all anteroposterior pelvic radiographs available on June 1, 2014, were retrospectively collected in a DICOM format. Additional information regarding sex, genotype, age at HSCT, enzyme replacement therapy pre-HSCT, IDUA activity post-HSCT, and donor chimerism was collected.

A set of radiographic parameters was constructed on the basis of a review of the literature, expert opinion, and feasibility and was reviewed by a group of experts. Each radiograph was evaluated by two independent observers using a strict protocol facilitated by OrthoGon software. The observers were trained by analysing between ten and twenty radiographs of patients not included in the study, following the study protocol. The training results were discussed, and where necessary, the definition of radiographic landmarks was modified to increase reproducibility. The observers were blinded to the patient's clinical data at the time of the evaluation.

Data analysis

First, interobserver variability was assessed by calculating the intraclass correlation coefficient (ICC, absolute agreement) on the basis of components of a variance analysis of a random selection of one radiograph per patient. For intraobserver variability, the ICC was calculated on the basis of ten images scored twice by each observer. Systematic errors were assessed by mixed-effects models, analysing the interaction of the average measurement of the two observers and the difference of the two measurements. Subsequently, Bland- Altman plots were constructed to identify outliers. Outliers were defined as data points outside the 95% limits of agreement. When a radiograph had one or more outliers, a determination was made whether to re-evaluate the complete radiograph, to exclude measurements from the final analysis of trends, or to include all findings despite the observed disagreement. In the final analyses, the average of the measurements of the two observers was used.

Reference values were obtained from the literature, and dichotomous variables were constructed for all parameters (normal or abnormal). The correlation of radiographic parameters was assessed on the basis of the last radiograph before the end of follow-up and calculated by the Pearson correlation coefficient. Binary outcomes were compared using the Fisher exact test.

Trajectories of individual patients were plotted over age for each parameter. Where applicable, average trends were fitted using mixed-effects models. Inter-individual variation was allowed for via random intercepts and slopes, and interaction with age at HSCT, enzyme replacement therapy prior to HSCT, donor chimerism, and IDUA activity post-HSCT was assessed. A p value of <0.05 was considered significant.

Demographic characteristics

Characteristic	Population (n =52)
Age at diagnosis	NR
Age at HSCT	
Mean age, months (std. dev.)	13.3 (7.2)
Median age, months (range)	12 (3–30)
≤18 months, n	NR
>18 months, n	NR
Gender, n (%)	

Study Reference	Langereis 2016	
	Male	30 (58)
	Female	22 (42)
	Race, n (%)	NR
	MPS I syndrome, n (%)	
	Hurler syndrome	52 (100)
	Scheie syndrome	0
	Hurler-Scheie syndrome	0
	Previous treatment	
	Patients who had received ERT, n (%)	31 (60)
	Mean duration of ERT pre-HSCT, weeks (SD)	24 (26)
	Median duration of ERT pre-HSCT, weeks (range)	14 (4–136)
	Iduronidase level	NR
	Mean number of radiographs per patient, (SD)	3.7 (1.8)
	Median number of radiographs per patient, (range)	3 (2–9)
Intervention	Intervention and comparator Haematopoietic stem cell transplantation. No distinct comparison groups. The correlation of radiographic findings with predictor variables was assessed.	
Outcomes Measured	Primary outcome The final set of parameters assessed from radiographs for the left and right hips consisted of the acetabular index, the neck-shaft angle, the pelvic tilt index, the migration percentage, the Smith ratio of lateral displacement, and the Smith ratio of superior displacement. The symphysis-os ischium angle and rotation quotient were also measured. Secondary outcomes Interaction between the radiographic parameters and potential predictors, such as age at transplantation, IDUA activity post-HSCT, donor chimerism, enzyme replacement therapy treatment pre-HSCT, or sex. The ICC for interobserver variation was also measured.	
Effectiveness of the Intervention	Efficacy No significant interaction was found between the radiographic parameters and potential predictors, such as age at transplantation. Related subgroup analysis (if applicable): NA Safety NR	
Authors' Conclusions	The study failed to identify any correlations between radiographic parameters and clinical characteristics, including age at HSCT. The study presents an extensive and reliable evaluation of radiographic parameters in a large cohort of patients with MPS I-H who received a successful HSCT. Most of the parameters studied showed a progressive deviation from reference values with age.	

Abbreviations: DICOM: Digital Imaging and Communications in Medicine; HSCT: haematopoietic stem cell transplantation; ICC: intraclass correlation coefficient; IDUA: α-L-iduronidase; MPS IH: mucopolysaccharidosis type I Hurler phenotype; NA: not applicable; NR: not reported.

Study Reference	Laraway 2016
Study Design	Design Retrospective chart review Objective To evaluate long-term outcomes of laronidase enzyme replacement therapy in patients with attenuated mucopolysaccharidosis type I (MPS I) Dates 2000 to 2009 Country UK Setting Department of Genetic Medicine, St Mary's Hospital, Manchester
Population Characteristics	Patient recruitment/eligibility Clinical data were reviewed retrospectively for living and deceased patients with a diagnosis of MPS I classified clinically and/or by molecular analysis as having attenuated disease, who started laronidase treatment during the study period at the study institution, and who did not undergo hematopoietic stem cell transplantation. Diagnosis of MPS I Classified clinically and/or by molecular analysis as having attenuated disease Duration of follow-up Mean follow-up of 6.1 years, median 6 years (range 1–10). Duration of follow-up varied for each outcome. Sample size N screened/invited = NR N eligible = 35 N excluded (with reason) = NR N lost to follow-up = NR N completed = NR N excluded from analysis = NR N included in analysis = 35 Data collection Case notes, laboratory results, and data from clinical trials were retrospectively reviewed. For each patient, the last data collected before ERT initiation were used as the baseline for clinical outcomes measured. Data analysis For uGAG, height-for-age Z scores, % predicated FVC, and 6MWT endpoints, a repeated-measures mixed model was applied. The covariates included age group (3 categories: ≤4, 5-9, and ≥10 years at treatment initiation), sex, ethnicity, duration on therapy, and a duration on therapy × age group interaction effect, with a patient-level random slope and intercept effect. Baseline value also was included as a covariate for the height-for-age Z scores, % predicted FVC, and 6MWT endpoints. Linear slope estimates from mixed models are from baseline through patients' follow-up. The repeated measures mixed model analysis was based on observed data; missing data were not imputed. Outcomes were considered statistically significant if the P-

value was ≤0.05. Aortic valve, mitral valve, corneal clouding, and visual acuity shift analyses for age categories present the patient's change in score from baseline using last observation carried forward methodology (using the last post-baseline value). For each endpoint, the shift analysis classifies a patient's last post-baseline value as having worsened relative to baseline (deteriorated), remained the same as baseline (stable), or improved relative to baseline (improved). Results are presented overall and stratified by baseline age.

Demographic characteristics

Characteristic	N=35
Age at presentation, years	
Median (range)	4 (0.25–9)
Mean	3.85
Age at baseline (treatment initiation), years	
Median (range)	11.3 (0.5–23.1)
Mean	11.5
≤4, n	9
5–9, n	5
≥10, n	21
Gender, n	
Male	13
Female	22
Race, n (%)	
European origin	22
Asian origin (Pakistani)	12
Uncertain	1
MPS I syndrome, n (%)	
Hurler syndrome	NA
Attenuated[a]	35 (100)

[a] Scheie or Hurler-Scheie not specified

Intervention

Intervention and comparator
ERT: Laronidase intravenous infusions at 100 U/kg body weight weekly. Outcomes were compared according to patient age treatment initiation (before or after 10 years of age)

Outcomes Measured

Primary outcome

Urinary glycosaminoglycan (uGAG) excretion, 6-minute walk test (6MWT), forced vital capacity (FVC), height-for-age Z score, cardiac status (measured by left ventricular function and aortic and mitral valve function), corneal clouding, and visual acuity

Secondary outcomes

Additional investigations included testing for anti-laronidase antibodies and reduced cellular uptake of laronidase

Effectiveness of the Intervention

Efficacy

- Mitral valve: Fewer children aged <10 years at treatment initiation experienced deterioration compared with patients aged ≥10 years at treatment initiation (14% vs 45% at the last assessment)
- Aortic valve: Fewer children aged <10 years at treatment initiation experienced aortic valve deterioration compared with patients aged 10 years (14% vs 40%)
- Corneal clouding: Of children aged <10 years at treatment initiation, 9% deteriorated compared with 25% of patients aged ≥10 years
- Visual acuity: A greater percentage of children aged <10 years deteriorated compared with patients aged ≥10 years (40% vs 14%)

Clinical measures	Age at treatment initiation, years	At last follow-up (1–10 years)		
		Deteriorated, n (%)	Stable, n (%)	Improved, n (%)
Mitral valve	Overall (n=34)	11 (32)	22 (65)	1 (3)
	≤4 (n=9)	2 (22)	7 (78)	0
	5–9 (n=5)	0	5 (100)	0
	Total <10 (n=14)	2 (14)	12 (86)	0
	Total ≥10 (n=20)	9 (45)	10 (50)	1 (5)
Aortic valve	Overall (n=34)	10 (29)	22 (65)	2 (6)
	≤4 (n=9)	2 (22)	7 (78)	0
	5–9 (n=5)	0	4 (80)	1 (20)
	Total <10 (n=14)	2 (14)	11 (79)	1 (7)
	Total ≥10 (n=20)	8 (40)	11 (55)	1 (5)
Corneal clouding	Overall (n=23)	4 (17)	18 (78)	1 (4)
	≤4 (n=8)	1 (13)	6 (75)	1 (13)
	5–9 (n=3)	0	3 (100)	0
	Total <10 (n=11)	1 (9)	9 (82)	1 (9)
	Total ≥10 (n=12)	3 (25)	9 (75)	0
Visual acuity	Overall (n=24)	6 (25)	8 (33)	10 (42)
	≤4 (n=7)	4 (57)	0	3 (43)
	5–9 (n=3)	0	2 (67)	1 (33)

	Total <10 (n=10)	4 (40)	2 (20)	4 (40)
	Total ≥10 (n=14)	2 (14)	6 (43)	6 (43)

Safety
NR

Authors' Conclusions	Laronidase treatment resulted in disease stabilization in the majority of patients with a mean follow-up of 6.1 years. Data suggest that early treatment may result in better outcomes. Three of the 35 patients in the study died during the review period; 2 of cardiac complications (1 following surgery for mitral valve replacement) and 1 of sudden death (which may also be related to cardiac involvement). All 3 were older than 10 years of age at the start of ERT, had well-established disease at the time of ERT, and received treatment for 3-6 years. These outcomes may support early initiation of treatment. Better outcomes in children treated before 10 years of age were also suggested by shift analyses for mitral valve, aortic valve, and corneal clouding. For visual acuity, more deteriorations were observed in children <10 years compared with children >10 years at treatment initiation, although improvements occurred at similar rates for both age groups.

Abbreviations: 6MWT: 6-minute walk test; ERT: enzyme replacement therapy; FVC: forced vital capacity; GAG: glycosaminoglycan. HSCT: haematopoietic stem cell transplantation; MPS: mucopolysaccharidosis; NR: not reported; uGAG: urinary glycosaminoglycan.

Study Reference	**Megens 2014**
Study Design	Design Retrospective cohort study Objective To assess the incidence of perioperative complications in children with MPS and the impact of enzyme replacement therapy (ERT) followed by hematopoietic stem cell transplantation (HSCT). Dates February 2003 to June 2012 Country The Netherlands Setting The national referral centre for HSCT at the Wilhelmina Children's Hospital
Population Characteristics	Patient recruitment/eligibility Medical records of patients with MPS, who received anaesthesia from February 2003 to June 2012, were reviewed. Patients with MPS treated with ERT prior to HSCT were included; patients treated with HSCT only, ERT only, HSCT followed by ERT or no treatment, were excluded from this study. Diagnosis of MPS I NR

Study Reference **Megens 2014**

Duration of follow-up

NR

Sample size
N screened/invited = 26
N eligible = 26
N excluded (with reason) = 7 (they were not treated according to the recent consensus strategy: ERT followed by HSCT [1 MPSI patient])

N lost to follow-up = 0
N completed = 19 (17 MPS I [Hurler])
N excluded from analysis = 0
N included in analysis = 19

Data collection
Preoperative details on physical health were obtained from the medical records. Perioperative anaesthesia data were retrieved from the Anesthesia Information and Management System. Perioperative respiratory complications were defined as any airway difficulty reported in the anaesthesia information and management system including difficult mask ventilation or use of an oropharyngeal airway, difficulty creating an unobstructed airway within the laryngeal mask airway, any change in intubation technique used, peripheral oxygen saturation below 90% for at least one minute, airway obstruction at emergence necessitating intervention, and/or unplanned admittance to the paediatric intensive care unit.

Data analysis

Data are presented in median with interquartile range or percentages with 95% confidence interval. Categorical variables were analysed using chi-square. Univariate and multivariate logistic regression analyses were performed to assess the association between the presence of airway problems at induction (outcome) and age (in months) and time after initiation of therapy (in months) as continuous data with generalised estimation equation adjustment for correlated records.

Demographic characteristics

Characteristic	Value
Age at diagnosis	NR
Age at intervention (ERT)	
Median age, years (range, months)	14 (7–43)
Age at intervention (HSCT)	
Median age, years (range, months)	18 (10–43)
Duration of ERT before HSCT started	
Median, months (range)	1 (0–15)
Gender, n (%)	NR
Race, n (%)	NR
MPS I syndrome, n	17
Hurler syndrome, n	17
Iduronidase level	NR

Intervention

Intervention and comparator
Intervention: ERT prior to HSCT; Comparator: outcome correlation with age

Study Reference	Megens 2014
Outcomes Measured	Primary outcome The incidence of perioperative respiratory and cardiovascular complications. Secondary outcomes The relationship between the phase of treatment by ERT followed by HSCT and patient age on the incidence of respiratory complications.
Effectiveness of the Intervention	Efficacy Univariate and multivariate logistic regression analyses did not show a relation between the incidence of airway management difficulty and age (OR: 1.01, P = 0.36). Related subgroup analysis (if applicable): NA Safety NR
Authors' Conclusions	Perioperative airway management was most successful using a laryngeal mask airway or video laryngoscope. Treatment with ERT followed by HSCT and patient age did not influence the incidence of perioperative respiratory problems.

Abbreviations: ERT: enzyme replacement therapy; HSCT: haematopoietic stem cell transplantation; MPS I: mucopolysaccharidosis type I; NA: not applicable; NR: not reported.

Study Reference	Poe 2014
Study Design	Design Prospective cohort study Objective To determine whether age at transplantation can predict cognitive outcomes in patients with Hurler syndrome Dates June 1997 to February 2013 Country USA Setting NR
Population Characteristics	Patient recruitment/eligibility Patients with Hurler syndrome who were referred to the Program for the Study of Neurodevelopment in Rare Disorders and subsequently underwent umbilical cord blood transplantation at other institutions during the study period. Diagnosis of MPS I

Study Reference	Poe 2014

For all patients, diagnosis of the more severe Hurler phenotype of mucopolysaccharidosis I was confirmed by clinical phenotype, including evidence of central nervous system involvement, and low IDUA enzyme levels in peripheral blood leukocytes.

Duration of follow-up
Median follow-up of 7.3 years (range 2 to 21.7), with a median of 7.0 evaluations (range 3 to 18).

Sample size
N screened/invited = 32
N eligible = 32
N excluded (with reason) = 1 (declined to participate)

N lost to follow-up = 0
N completed = 31
N excluded from analysis = 0
N included in analysis = 31

Data collection

- All patients were evaluated on the same day by a neurodevelopmental paediatrician working with audiologists, speech therapists, psychologists, and physical therapists at baseline and every 6 to 12 months after transplantation
- Physical, neurological, and visual and hearing examinations were performed at each evaluation, including assessment of corneal clouding and documentation of use of eyeglasses as a marker of decreased visual acuity
- Audiological function was assessed by behavioural audiometry for children older than 6 months, and by otoacoustic emissions for younger children. Tympanometry was performed in all patients
- Brainstem auditory evoked potentials were used to assess the degree of sensorineural hearing loss in patients who were difficult to assess using behavioural audiometry and those for whom further information regarding auditory brainstem function was needed
- Standardised and validated neurobehavioral tools were used to assess all children. Cognitive, adaptive, and language function were longitudinally assessed

Data analysis

- To evaluate disease progression, the baseline cognitive score was used to create a developmental quotient (developmental age/calendar age). Developmental curves were generated and analysed to determine the effect of disease progression and age at transplantation on outcomes. The resulting developmental trajectories were compared to norms of typically developing children
- The relationship of age at transplantation and baseline cognitive ratio with post-transplant developmental trajectories was evaluated using general linear mixed models. Independent variables were age at the time of testing, age, baseline developmental quotient, age at transplantation (months), and the interactions between developmental quotient and age at testing, and age at transplantation and age at testing. Developmental score was the dependent variable. The models assumed random intercept and slope (age), which was centred at 0 months of age. To aid interpretation, the interaction terms were evaluated and removed if nonsignificant ($p>0.10$). The regression coefficients (β) were evaluated for significance
- Although age was entered as a continuous variable, results are reported for groups based on age at transplantation to help visualise the effect of age

Study Reference **Poe 2014**

Demographic characteristics

Characteristic	Value
Age at diagnosis	NR
Age at transplantation	
Median age, months (range)	13.8 (2.1–34.3)
Gender, n	
Male	16
Female	15
Race, n	
White	29
Black	2
MPS I syndrome, n (%)	
Hurler syndrome	31 (100)
Scheie syndrome	NA
Hurler-Scheie syndrome	NA
Previous treatment	
Patients who had received ERT, n	0
Iduronidase level	NR
Median developmental quotient (range)	0.86 (0.49–1.34)

*and other baseline characteristics reported for individual cases only.

Intervention and comparison groups

Intervention

Umbilical cord blood transplantation.

Patients underwent conditioning with busulfan, cyclophosphamide, and horse antithymocyte globulin. Prophylaxis against graft-versus-host disease was carried out by using cyclosporine and methylprednisolone. Supportive care was administered in most patients. The majority of patients underwent tonsillectomy, adenoidectomy, and pressure-equalising tube placement before transplantation.

Comparison groups

Outcomes were analysed in three patient groups by age of transplantation

- Group 1: 2–8 months old at transplantation (n=6)
- Group 2: 9 to 17 months old at transplantation (n=17)
- Group 3: ≥18 months old at transplantation (n=8)

Outcomes Measured

- Neurodevelopmental function. A standardised neurodevelopmental protocol was used to assess patients who were referred for pre- and post-transplant evaluation
- Cognitive skills assessed using the Mullen Scales of Early Learning, and in later years, the Differential Ability Scales

Study Reference	Poe 2014
	• Adaptive behaviour assessed using the Scales of Independent Behavior-Revised • Language skills assessed using the Preschool Language Scale and the Clinical Evaluation of Language Fundamentals • Motor skills assessed using the Peabody Developmental Motor Scales • Audiological and visual function (No outcomes were specified as primary or secondary)
Effectiveness of the Intervention	**Efficacy** Cognitive function Age at transplantation was a strong predictor of post-transplant cognitive development (p<0.001). Younger age at transplantation was associated with greater gains in cognitive function during the follow up period (β: −0.024, p<0.001). The interaction between baseline cognitive score and age was not significant (p=0.34). For example, at the calendar age of 12 years, the average developmental age is expected to be 12.3 years if the patient was transplanted at 4 months, 10.3 years if transplanted at 12 months, and 7.0 years if transplanted at 25 months. Thus, children transplanted at 12 and 25 months of age display cognitive functioning at a level 2 to 5.3 years below that of children transplanted at 4 months. Adaptive behaviour Younger age at transplantation was associated with better post-transplant adaptive behaviour development (β: −0.013, p=0.030). The rate of acquisition decreased over time (as shown by the negative estimate for the age^2 coefficient (β: −0.026, p<0.001). Baseline cognitive ratio was not significantly related to adaptive function (p=0.07). Language skills Children who underwent transplantation at younger ages had better skill development than those transplanted later (receptive language, β: −0.022, p=0.004; expressive language, β: −0.023, p=0.010). Audiological and visual function • Severity of hearing loss and use of hearing aids were similar across patient groups, regardless of age at transplantation • The severity of corneal clouding and rates of improvement, stabilisation and worsening after transplantation did not differ across groups according to age at transplantation (n=28) • Rates of corneal transplantation and use of eyeglasses were similar across the 3 groups, and all patients had adequate visual acuity to perform cognitive testing Related subgroup analysis: NR Safety NR

Study Reference	Poe 2014
Authors' Conclusions	We show that age at transplantation is a strong predictor of long-term cognitive and language outcomes in children with Hurler syndrome. Our results emphasise the urgent need for early identification and treatment of Hurler syndrome, implementation of newborn screening for this disease, and identification of early markers of neurological disease.

Abbreviations: ERT: enzyme replacement therapy; IDUA: α-L-iduronidase; MPS: mucopolysaccharidosis; NR: not reported.

Study Reference	Pal 2015
Study Design	Design Retrospective cohort study Objective To describe the pattern of sleep disordered breathing (SDB) seen in the largest MPS I cohort described to date and determine therapies and biomarkers influencing the severity of long-term airway disease. Dates 2003 to 2013 Country UK and The Netherlands Setting Royal Manchester Children's Hospital, Manchester and the Academic Medical Centre, Amsterdam
Population Characteristics	Patient recruitment/eligibility A study group of 61 patients in whom long-term follow-up was available at the 2 centres was included in the study. Patients were recruited, following consent to donate blood, which is routinely sought from all MPS patients in Manchester. Diagnosis of MPS I 44 patients were classified as severe Hurler based on genotype and enzyme studies at diagnosis, while 17 were diagnosed as attenuated phenotype Duration of follow-up The median duration of follow-up was 22 months (range: 1–60 months). Sample size N screened/invited = NR N eligible = 61 N excluded (with reason) = NR N lost to follow-up = NR N completed = NR N excluded from analysis = NR N included in analysis = 61

Data collection

The clinical notes of patients were retrospectively reviewed for over-night sleep oximetry studies, therapeutic and biochemical data.

Data analysis

Significant associations between patient, biomarker and treatment-related variables and primary endpoints for all patients, HSCT treated Hurler and ERT treated attenuated subgroups were identified using multivariate stepwise regression modelling. Possible confounding variables including age at start of treatment and duration of follow-up were included in the model. Significance was assumed where 95% confidence intervals did not include 1 (p-values <0.05).

Demographic characteristics

Characteristic	Value
All MPS I	
Age at diagnosis	NR
Age at start of treatment, months, median (range)	18 (3–364)
Age at final assessment, months, median	82 (0.3–420)
Gender, n (%)	
Male	38 (62)
Female	23 (38)
Race, n (%)	NR
MPS I syndrome, n (%)	
Hurler syndrome	44 (69)
Scheie syndrome	NR
Hurler-Scheie syndrome	NR
Attenuated	17 (31)
HSCT treated Hurler patient characteristics	
Age at start of treatment (months), median (range)	14 (3–30)
Number of HSCT, n (%)	
1	32 (78)
2	8 (20)
3	1 (2)
Source, n (%)	
Cord blood	17 (41)

Study Reference	Pal 2015
	Bone marrow: 15 (37) Peripheral blood stem cells: 7 (17) Unknown: 2 (5) Donor, n (%) Related: 15 (37) Matched unrelated donor: 26 (63) **ERT treated Hurler patient characteristics** Age at start of treatment (months), median (range): 85 (74–144) **ERT treated attenuated patients** Age at start of treatment (months), median (range): 60 (24–364)
Intervention	Intervention and comparator Patients were not divided into intervention groups for the purpose of the analysis; instead, significant associations between patient, biomarker and treatment-related variables and primary endpoints were identified for all patients (N=61), and in the HSCT treated Hurler (N=44) and ERT treated attenuated subgroups (N=17).
Outcomes Measured	Primary outcome The primary endpoint measured was the presence, progression and severity of sleep disordered breathing (SDB), based on overnight sleep oximetry data (the effect of age at treatment initiation was not explored for primary outcomes). Secondary outcomes Leukocyte IDUA enzyme activity one-year post HSCT, the urine dermatan sulphate: chondroitin sulphate (DS:CS) ratio biomarker and requirement for therapeutic airway intervention following commencement of treatment for MPS (the only outcome for which the correlation with age at treatment was reported).
Effectiveness of the Intervention	Efficacy Sixteen percent (10/61) of the study population required 13 episodes of therapeutic airway intervention following initiation of ERT/HSCT, with a later age at start of treatment found to significantly correlate with the need for intervention (p=0.012). Safety NR
Authors' Conclusions	Interventions maximising substrate reduction correlate with improved long-term SDB, while inhibitory antibodies impact on biochemical and clinical outcomes. Monitoring and tolerisation strategies should be re-evaluated to improve detection and minimise the inhibitory antibody response to ERT in MPS I and other lysosomal storage diseases. Future studies should consider the use of SDB as an objective parameter of clinical and metabolic improvement.

Abbreviations: DS:CS: dermatan sulphate: chondroitin sulphate; ERT: enzyme replacement therapy; HSCT: haematopoietic stem cell transplantation; IDUA: α-L-iduronidase; MPS: mucopolysaccharidosis; NR: not reported; SDB: sleep disordered breathing.

Study Reference	Rodgers 2017
Study Design	Design Retrospective study Objective To determine how hematopoietic stem cell transplantation (HSCT) has altered mortality in an institution's 30-year experience of patients with MPS IH undergoing HSCT. Dates September 1983 to December 2013 Country US Setting University of Minnesota
Population Characteristics	Patient recruitment/eligibility The University of Minnesota Division of Pediatric Blood and Marrow Transplantation has maintained a registry of all patients who underwent HSCT and updates patient vital status annually. All patients with MPS IH who underwent HSCT between the program's inception and 12/31/2013 were included in this study. Diagnosis of MPS I Early in the program the diagnosis of Hurler syndrome was established by characteristic phenotype and presentation within the first 2 years of life and verified by measurement of leukocyte IDUA. In later years the diagnosis was confirmed by mutation analysis. Duration of follow-up The median length of follow-up after final transplant was 10.7 (IQR 5.0–17.2), with a maximum follow-up of 28.97 years. Sample size N screened/invited = 134 N eligible = 122 (vital status confirmed) N excluded (with reason) = 12 (censored at the last date of contact) N lost to follow-up = NR N completed = NR N excluded from analysis = NR N included in analysis = 134 Data collection Within the study institution's registry, patient vital status is categorized as alive, deceased, and unknown with last date of follow-up. Known decedents and patients with unknown vital status were cross-referenced with the CDC's NDI, a centralized data repository containing US death records from 1983 through 2013, to confirm vital status and, if deceased, date of death and proximate and underlying causes of death.

Study Reference **Rodgers 2017**

Publicly available social media and results from an unrelated IRB-approved study were used to confirm vital status in those remaining unknown. Causes of death for known decedents were verified through retrospective medical chart review and classified into one of four major categories: cardiac, pulmonary, infectious disease, and other (which included haematologic/oncologic, neurologic, gastrointestinal or unknown). NDI proximate cause of death was verified by patient's clinical data, including documentation leading up to the death, death note, and autopsy findings, if performed.

Data analysis

Time at risk for survival analyses began at the time of successful or final transplantation until death or censoring (lost-to-follow-up or administratively on Dec 31 2013). Unadjusted survival curves were based on Kaplan-Meier estimates. Adjusted analyses used Cox proportional hazards regression models adjusting for sex, transplant era, age categories at transplantation (<12 months, 12–24 months, and >24 months) and percentage engraftment categories (≥90%, 10–89%, and <10%) with robust variance estimation for confidence intervals and p values. Cause of death over time was evaluated using cumulative incidence curves to account for potential competing risks.

Demographic characteristics

Characteristic	All patients (N=134)
Male, n (%)	69 (51.5)
Race, n (%)	NR
MPS I syndrome, n (%)	
Hurler syndrome	134 (100)
Scheie syndrome	NA
Hurler-Scheie syndrome	NA
Age at diagnosis	NR
Mean age at HSCT, months (SD)	21.8 (20.8)
Age at HSCT, n (%)	
<12 months	26 (19.4)
12 to 24 months	75 (56.0)
>24 months	33 (24.6)
Number of HSCTs	
One	112 (83.6)
Two	21 (15.7)
Three	1 (0.7)
HSCT type	
Marrow related	35 (26.1)
Marrow unrelated	48 (35.8)
Unrelated UCB	49 (36.6)
Peripheral blood stem cell	2 (1.5)
Other treatment	
Peri-HSCT ERT	38 (28.4)

Study Reference	**Rodgers 2017**
	No ERT: 96 (71.6)
Intervention	Intervention: Allogenic HSCT Comparator: The entire cohort was separated into transplant era subgroups (prior to 1st January 2004 versus 2004 and beyond) and sex subgroups (male vs female), and these subgroups were compared against each other. Patients were stratified according to age at HSCT (<12 months, 12 to 24 months and >24 months) and these groups were compared against each other.
Outcomes Measured	Primary outcome The primary endpoint was survival including, if available, cause of death after final allogeneic HSCT. Secondary outcomes Other patient and treatment characteristics measured included sex and age at first successful HSCT, number and type of HSCT, use of any total body or total lymphoid irradiation (TBI/TLI) or targeted busulfan dosing as part of the preparative phase, use of peri-HSCT laronidase ERT, most recent known percentage donor myeloid hematopoietic chimerism, leukocyte IDUA level for the subgroup of those surviving >1 year after HSCT, and vital status at the end of 2013.
Effectiveness of the Intervention	Efficacy Cox proportional hazards model results for survival over the first 8 years post-HSCT

Covariate	Hazard ratio	95% CI	p value
Age at treatment <12 month	Ref	-	-
Age at treatment 12–24 months	1.42	0.59–3.40	0.429
Age at treatment >24 months	1.45	0.52–4.00	0.475

Effectiveness of the Intervention	Safety NR
Authors' Conclusions	The authors did not refer to the finding that there is no significant difference in survival based on age at treatment being <12 months, 12–24 months or >24 months.

Abbreviations: CDC: Centers for Disease Control; CI: confidence interval; ERT: enzyme replacement therapy; HSCT: haematopoietic stem cell transplantation; IDUA: alpha-L-iduronidase; IRB: institutional review board; IQR: inter-quartile range; MPS IH: mucopolysaccharidosis I, Hurler syndrome; NDI: National Death Index; NR: not reported; SD: standard deviation; TBI: total body irradiation; TLI: total lymphoid irradiation; UCB: umbilical cord blood; US: United States.

Study Reference	Wadhwa 2018
Study Design	Design Retrospective study Objective To provide a comprehensive assessment of overall and cause-specific late mortality in patients with specific types of inborn errors of metabolism (IEM) treated with allogeneic BMT, to inform both the transplantation physicians as well as the patients and their families. Dates 1974 to 2014 Country United States Setting City of Hope, University of Minnesota and University of Alabama at Birmingham
Population Characteristics	Patient recruitment/eligibility Patients who had undergone allogenic BMT during the study period for a diagnosis of IEM and had survived for ≥2 years after BMT were included. Diagnosis of MPS I NR Duration of follow-up Median follow-up of 13.2 years (range 1.6–31.4 years) Sample size N screened/invited = NR N eligible = 273 N excluded (with reason) = 0 N lost to follow-up = 0 N completed = 273 N excluded from analysis = 0 N included in analysis = 273 (96 Hurler [MPS I] patients) Data collection Information on demographic characteristics, primary diagnosis, preparative regimens, stem cell source, type of donor (sibling versus unrelated), agents used for graft-versus-host disease (GVHD) prophylaxis and presence of chronic GVHD was obtained from the institutional transplantation databases and supplemented by medical records. Chronic GVHD was classified as none, limited, or extensive based on the Seattle Classification. NDI Plus and/or medical records provided information regarding the date and cause of death through December 2015. Additional information from the Accurint database was used to extend the vital status information through December 2016. All patients were assigned a primary and, if present, secondary cause of death, independently by 2 investigators. Cause of death assignments were further verified by a third investigator and in the event of discrepant assignments, a fourth investigator provided adjudication. Causes of death (primary or secondary) attributable to the underlying IEM were classified as disease-related mortality. Causes of death due to therapeutic exposures and the transplantation procedure were collectively classified as transplantation-related mortality.

Study Reference **Wadhwa 2018**

Data analysis

Kaplan-Meier techniques were used to describe overall survival. The cumulative incidence of cause-specific mortality was calculated using competing-risk methods. The standardized mortality ratio (SMR), a ratio of observed to expected number of deaths, was used to compare the mortality experienced by this cohort to the age- (5-year intervals), sex-, and calendar-specific (5- year intervals) mortality of the general US population, using data obtained from the Centres for Disease Control and Prevention. The 95% confidence intervals (CIs) of the SMR were calculated using the Poisson regression method. SMRs were calculated for the entire cohort, as well as separately by sex, year of transplantation (1983 to 1988, 1989 to 1994, 1995 to 2000, 2001 to 2006, and 2007 to 2014), primary IEM diagnosis, and presence or absence of chronic GVHD. Cox regression analysis was used to identify predictors of all-cause mortality for the entire cohort, as well as for the most prevalent IEM diagnoses considered individually. Owing to the small number of subjects and deaths in each model and the associated collinearity among the variables, a parsimonious model was created using the variables with associated P values <0.1 in the multivariable model. Furthermore, owing to the varied clinical presentations of different IEM and the significant differences in median age at BMT by type of IEM, the median age at BMT of each primary disease was used to evaluate its effect on late mortality.

Demographic characteristics

Characteristic	Value
Patients with IEM diagnosis, n	273
Hurler Syndrome, n (%)	96 (35.2)
Hurler Syndrome Patients	
Age at diagnosis	NR
Median age at intervention, years (range)	1.5 (0.4–6.0)
Median time since intervention, years (range)	13.2 (1.6–31.4)
Sex, n (%)	
Male	57 (59.4)
Female	39 (40.6)
Race, n (%)	
Non-Hispanic white	93 (96.9)
Hispanic	2 (2.1)
Black	1 (1.0)
Type of donor, n (%)	
Sibling	24 (25)
Unrelated	70 (72.1)
Parent	2 (2.1)
Stem cell source, n (%)	
Bone marrow	53 (55.2)
Cord blood	43 (44.8)
Peripheral blood stem cells	0
Previous treatment	
Patients who had received ERT, n	NR
Iduronidase level	NR
Therapies used in the preparative regime, n (%)	

Study Reference	Wadhwa 2018
	Total body irradiation: 31 (32.3) *Cyclophosphamide*: 86 (89.6) *Busulfan*: 83 (86.5) Agent used for GVHD prophylaxis, n (%) *Cyclosporine*: 86 (89.6) *Mycophenolate mofetil*: 29 (30.2) *T cell depletion*: 25 (26.0)
Intervention	Intervention and comparator Intervention: Allogeneic BMT. Age- and sex- matched cohort from the general US population. Sub-analyses compared the group receiving BMT at <the median age of intervention (1.5 years) with the group receiving BMT at ≥the median age of intervention.
Outcomes Measured	Primary outcome • Survival rate • Age at death • Risk of premature death compared to age-, sex- and calendar-specific rates observed in the general US population • Disease-related mortality Secondary outcomes Outcomes were not split into primary and secondary outcomes.
Effectiveness of the Intervention	Efficacy Age at BMT had a significant impact on survival in patients with Hurler syndrome. Early BMT at less than the median age at BMT of 1.5 years, was associated with a lower risk of all-cause late mortality (HR: 0.2; 95% CI: 0.05 to 0.9; P = 0.03) compared with BMT performed at older than the median age. Related subgroup analysis (if applicable): NA Safety Safety outcomes were not reported in relation to age at BMT. The majority of the causes of death for Hurler syndrome (n=14) were attributable to the primary disease (n=10). Other causes of death included infection (n=1), subsequent malignant neoplasms (n=1), cardiac disease (n=1) and pulmonary disease (n=1)
Authors' Conclusions	The data show that patients with IEM who undergo allogeneic BMT and survive for at least 2 years have a relatively favourable overall survival even at 20 years after BMT. Younger age at BMT and use of busulfan and cyclosporine were protective in patients with Hurler syndrome.

Abbreviations: BMT: blood or marrow transplantation; CI: confidence interval; ERT: enzyme replacement therapy; HSCT: haematopoietic stem cell transplantation; IDUA: alpha-L-iduronidase; IQR: inter-quartile range; MPS IH: mucopolysaccharidosis I, Hurler syndrome; NDI: National Death Index; NR: not reported; SD: standard deviation; US: United States.

Study Reference	Wyffels 2017
Study Design	Design Retrospective chart review Objective The primary objective was to determine if the use of limited ERT in addition to HCT for the treatment of children with HS reduces the incidence of surgical intervention for CTS compared with a cohort of historical controls treated with HCT alone. The secondary objectives were to evaluate the impact of demographic and transplant-related characteristics on the incidence of CTS and the results of surgical treatment of CTS in HS. Dates 1985 to 2012 Country US Setting Two participating hospitals (NR)
Population Characteristics	Patient recruitment/eligibility Medical records from 2 participating hospitals in Minnesota were reviewed for patients with a diagnosis of HS during the study period. Patients were eligible if they had a diagnosis of mucopolysaccharidosis type 1H (HS), were treated with HCT or HCT and ERT with at least one year survival, and were evaluated by hand surgery including nerve conduction study (NCS) testing. Diagnosis of MPS I NR Duration of follow-up Mean follow-up was 5 years for group 2 (HS children treated with HCT and ERT) and 15 years for group 1 (historical cohort undergoing HCT alone). Sample size N screened/invited = 83 N eligible = 74 (group 1: 43; group 2: 31) N excluded (with reason) = 9 (8 [not surviving 1 year after HCT] + 1 [no hand surgery evaluation]) N lost to follow-up = 0 N completed = 74 N excluded from analysis = 0 N included in analysis = 74 Data collection Demographics, transplant characteristics, and nerve conduction velocity (NCV) testing results were recorded for all patients. Between the ages of 2 and 5 years, all transplanted patients were evaluated by a hand surgeon along with a screening NCS. Data analysis

Study Reference **Wyffels 2017**

The frequency of CTS was normalised by comparing those who underwent surgery within 5 years of HCT (due to differences in the mean follow-up time between group 1 (15 years) and group 2 (5 years). The cumulative incidence function and two-sided log-rank test were used to evaluate time from HCT to CTS surgery. Two-sided unequal variance t-tests were used to compare NCS measurements between groups and from pre- to post-surgical state.

Demographic characteristics

Characteristic	Group 1: HCT (N=43)	Group 2: HCT + ERT (N=31
Age at diagnosis	NR	NR
Age at intervention		
Mean age, years (range)	1.8 (0.5–6.0)	1.5 (0.4–2.9)
Gender, n (%)		
Male	24 (56)	17 (55)
Female	NR	NR
Race, n (%)	NR	NR
MPS I syndrome, n (%)		
Hurler syndrome	43 (100)	31 (100)
Scheie syndrome	NA	NA
Hurler-Scheie syndrome	NA	NA
Related marrow donor, n (%)	19 (44)	6 (19)
Unrelated marrow donor, n (%)	19 (44)	2 (6)
Umbilical cord blood, n (%)	5 (12)	23 (74)
Previous treatment	NR	NR
Patients who had received ERT, n	NR	NR
Iduronidase level	NR	NR

Intervention

Intervention and comparator

Group 2 (intervention [n=31]): HCT + ERT; patients transplanted from 2003 to 2012 received peritransplant laronidase ERT administered intravenously at a dose of 0.58 mg/kg weekly. Patients in group 2 generally received 12 laronidase doses pre-transplant and 8 doses following transplant, although the actual total doses varied based on the clinical situation.

Group 1 (comparator [n=43]): HCT; patients transplanted prior to 2003 had been treated with HCT alone.

Outcomes Measured

Primary outcome

The incidence of CTS in HS children treated with HCT + ERT compared with a historical cohort under-going HCT alone. CTS diagnosed by nerve conduction speed measurements (pre- and post-surgery): A positive study for CTS would typically be concluded if the median/ulnar peak latency (sensory or motor) constituted a 200% increase, or the corresponding median/ulnar conduction velocity was slowed to 70% or less, as correlated to their clinical findings by the hand specialist.

Secondary outcomes

The effect of age at transplant, graft type, and sex on the incidence of CTS.

Study Reference	Wyffels 2017
Effectiveness of the Intervention	Efficacy No difference was established for a delay in HCT until after the age of 2 years (5-year CTS incidence for age < 2 years was 48% [95% CI, 32–62], for age > 2 years 47% [95% CI, 21–68]). Represented graphically. Related subgroup analysis (if applicable): NA Safety NR
Authors' Conclusions	In the current study, patients younger than 2 years old at transplant had the same incidence of carpal tunnel surgery as patients older than 2 years. Although the administration of ERT prior to and for several months after HCT has become routine in the authors' institution, the findings do not suggest this combined therapy is sufficient to decrease the development of CTS.

Abbreviations: A.V.H.: evaluation by hand surgery; ERT: enzyme replacement therapy; CTS: carpal tunnel syndrome; HCT: haematopoietic stem cell transplantation; HS: Hurler syndrome; NA: not applicable; NCS: nerve conduction study; NR: not reported; US: United States.

Appraisal for quality and risk of bias

Quality assessments of included studies are reported below.

Table 17. Guidance for QUADAS-2 quality assessment of studies extracted for question 1

Question	Guideline Criteria for MPS I Studies	Literature-Recommended Criteria
PARTICIPANT SELECTION		
Was a consecutive or random sample of newborns enrolled?	**Yes** if all newborns (or a random sample of patients) within the study period were included **No** if patients were selected in a different way, e.g. by referral or convenience sample **Unclear** if all screened newborns are enrolled but it is not specified if the screening test is routinely administered at the study site	A study should ideally enrol all consecutive, or a random sample of, eligible patients – otherwise there is potential for bias. Studies that make inappropriate exclusions, e.g. excluding "difficult to diagnose" patients, may result in overoptimistic estimates of diagnostic accuracy
Was a case-control design avoided?	**Yes** if the study was a prospective or retrospective cohort study **No** if cases of MPS I were matched to controls	Studies enrolling patients with known disease and a control group without the condition may exaggerate diagnostic accuracy
Did the study avoid inappropriate exclusions?	**Yes** if all newborns were included, or if exclusions were appropriate and unlikely to lead to bias **No** if any group within the screening population was systematically excluded **Unclear** if it is unclear if exclusions were made or exclusions were made but it is unclear if these were appropriate	Exclusion of patients with "red flags" for the target condition, who may be easier to diagnose, may lead to underestimation of diagnostic accuracy
Could the selection of newborns have introduced bias?	Answered based on the previous questions in this domain	If all signalling questions for a domain are answered "yes" then risk of bias can be judged "low". If any signalling question is answered "no" this flags the potential for bias
Is there concern that the included newborns do not match the review question?	**Low** if patients overall are newborn babies representative of the screening population (i.e. similar to the newborn population in the UK) **High** if patients overall are not representative of the screening population, such as newborns in families known to be affected by MPS I or demographically dissimilar to the UK population **Unclear** if it is unclear whether the population is similar to the UK newborn population	There may be concerns regarding applicability if patients included in the study differ, compared to those targeted by the review question, in terms of severity of the target condition, demographic features, presence of differential diagnosis or co-morbidity, setting of the study and previous testing protocols
INDEX TESTS		
Were the index test results interpreted without knowledge of the reference standard?	**Yes** if screening results were interpreted before the diagnosis was confirmed **No** if screening results were only examined after the diagnosis was confirmed **Unclear** if it is unclear if screening results were interpreted before or after the diagnosis was confirmed	This item is similar to "blinding" in intervention studies. Interpretation of index test results may be influenced by knowledge of the reference standard
If a threshold was used, was it pre-specified?	**Yes** if the criteria used to diagnose MPS I were explicitly stated, well-defined, and specified before the study **No** if criteria were not stated, were insufficiently well-defined, were specified retrospectively or adjusted during the study	Selecting the test threshold to optimise sensitivity and/or specificity may lead to overoptimistic estimates of test performance, which is likely to be poorer in an independent sample of patients in whom the same threshold is used

Could the conduct or interpretation of the index test have introduced bias?	Answered based on the previous questions in this domain. Consider whether the staff conducting the index test could have had foreknowledge of who was at risk by presence of major factors.	If all signalling questions for a domain are answered "yes" then risk of bias can be judged "low". If any signalling question is answered "no" this flags the potential for bias
Is there concern that the index test, its conduct, or interpretation differ from the review question?	**Low** if the timing and manner of the index test are similar to that of the heel prick test currently used in the UK newborn population as part of other screening programmes **High** if any aspect of the index test, including its conduct or interpretation, is substantially different from the heel prick test currently used in the UK newborn population as part of other screening programmes **Unclear** if it is unclear whether the index test is similar to the heel prick test	Variations in test technology, execution, or interpretation may affect estimates of its diagnostic accuracy. If index tests methods vary from those specified in the review question there may be concerns regarding applicability
REFERENCE STANDARD		
Is the reference standard likely to correctly classify the test condition?	**Yes** if MPS I was confirmed through tests that diagnose both phenotypic and genetic aspects of the severe and attenuated forms of the condition **No** if diagnosis was performed inconsistently, the methods used are likely to be unreliable or if the confirmation diagnosis was not specific for MPS I **Unclear** if it is unclear what the reference standard was or whether it was administered consistently	Estimates of test accuracy are based on the assumption that the reference standard is 100% sensitive and specific. Disagreements between the reference standard and index test are assumed to result from incorrect classification by the index test
Were the reference standard results interpreted without knowledge of the results of the index test?	**Yes** if the final diagnosis of MPS I was made by an investigator blinded to the index test results **No** if the screening results were known by the investigator making the final diagnosis **Unclear** if it is not clear whether the investigator was aware of the test result when making the final diagnosis	Potential for bias is related to the potential influence of prior knowledge on the interpretation of the reference standard
Could the reference standard, its conduct, or its interpretation have introduced bias?	Answered based on the previous questions in this domain	If all signalling questions for a domain are answered "yes" then risk of bias can be judged "low". If any signalling question is answered "no" this flags the potential for bias
Is there concern that the target condition as defined by the reference standard does not match the review question?	**Low** if the reference standard is expected to correctly diagnose the whole spectrum of severe and attenuated forms of the condition, excluding individuals with pseudodeficiency who are not expected to develop symptoms **High** if the reference standard is likely to only include a subset of the disease spectrum (e.g. Hurler syndrome only) or if the reference standard also includes individuals who may be asymptomatic (e.g. with pseudodeficiency) **Unclear** if it is unclear whether the reference standard is expected to correctly diagnose the condition	The reference standard may be free of bias but the target condition that it defines may differ from the target condition specified in the review question. For example, when defining urinary tract infection, the reference standard is generally based on specimen culture but the threshold above which a result is considered positive may vary
PARTICIPANT FLOW		
Did all participants receive a reference standard? **Did participants receive the same reference standard?**	**Yes** if all screened patients had confirmation of their diagnosis, and all were diagnosed in the same manner (using the same reference standard by similarly trained staff) **No** if patients received different reference standards **Unclear** if there was a high variability in staff diagnosing MPS I	Verification bias occurs when not all of the study group receive confirmation of the diagnosis by the same reference standard. If the results of the index test influence the decision on whether to perform the reference standard or which reference standard is used, estimated diagnostic accuracy may be biased
Were all newborns included in the analysis?	**Yes** if all screened babies were included in the final analysis **No** if any screened babies were not included in the final analysis **Unclear** if it is unclear is any screened babies were excluded from the final analysis	All patients who were recruited into the study should be included in the analysis. There is a potential for bias if the number of patients enrolled differs from the number of patients included in the 2x2 table of results, for example because patients lost to follow-up differ systematically from those who remain

Could the participant flow have introduced bias?	**No** if newborns who underwent the index test all had the same chance of being diagnosed positive/negative if they had/not had MPS I **Yes** if newborns received different reference standards or a significant proportion were removed from the analysis **Unclear** if it is unclear whether newborns all had the same chance of being diagnosed as having or not having MPS I	If all signalling questions for a domain are answered "yes" then risk of bias can be judged "low". If any signalling question is answered "no" this flags the potential for bias

Table 18. Quality assessment of studies included for question 1

Question	Burlina 2018	Chuang 2018	Hopkins 2015	Minter Baerg 2018
PARTICIPANT SELECTION				
Was a consecutive or random sample of newborns enrolled?	Yes	Yes	Yes	Yes
Was a case-control design avoided?	Yes	Yes	Yes	Yes
Did the study avoid inappropriate exclusions?	Yes	Yes	Yes	Yes
Could the selection of newborns have introduced bias?	Low	Low	Low	Low
Is there concern that the included newborns do not match the review question?	Low	High	Low	Low
INDEX TESTS				
Were the index test results interpreted without knowledge of the reference standard?	Yes	Yes	Yes	Yes
If a threshold was used, was it pre-specified?	Yes	Yes	Yes	No
Could the conduct or interpretation of the index test have introduced bias?	Low	Low	Low	Unclear
Is there concern that the index test, its conduct, or interpretation differ from the review question?	High	Unclear	Unclear	High
REFERENCE STANDARD				
Is the reference standard likely to correctly classify the test condition?	Yes	Yes	No	Yes
Were the reference standard results interpreted without knowledge of the results of the index test?	Unclear	Unclear	Unclear	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low	Low	High	Low
Is there concern that the target condition as defined by the reference standard does not match the review question?	Low	High	High	High
PARTICIPANT FLOW				
Did all participants receive a reference standard?	No	Yes	No	No
Did participants receive the same reference standard?	Yes	Yes	Unclear	Yes
Were all newborns included in the analysis?	Yes	Yes	No	Yes
Could the participant flow have introduced bias?	Yes	Yes	Yes	Yes

Table 19. Guidance for ROBINS-1 quality assessment of studies extracted for question 2

Question	Possible answer/reason for not using question	Guideline criteria
BIAS DUE TO CONFOUNDING		

Question	Possible answer/reason for not using question	Guideline criteria
1.1 Is there potential for confounding of the effect of intervention in this study?	Y/PY/PN/N (Yes/Probably yes/Probably no/No)	Factors likely to influence the effect of interventions: gender, age, age at diagnosis, timing of intervention, disease severity, donor type, HLA disparity, total nucleated cells infused, CD34+ cells infused, prophylaxis received
1.2 If Y/PY to 1.1: Was the analysis based on splitting participants' follow up time according to intervention received?	Y/PY/PN/N	If participants could switch between intervention groups then associations between intervention and outcome may be biased by time-varying confounding. This occurs when prognostic factors influence switches between intended interventions.
1.3 Were intervention discontinuations or switches likely to be related to factors that are prognostic for the outcome?	Y/PY/PN/N	If intervention switches are unrelated to the outcome, for example when the outcome is an unexpected harm, then time-varying confounding will not be present and only control for baseline confounding is required.
1.4 Did the authors use an appropriate analysis method that controlled for all the important confounding domains?	Not applicable(NA)/Y/PY/PN/N/No information (NI) If there were analyses to control for confounding variables, (assuming the same variables as listed in 1.1) answer should be Y or PY, depending on whether there were differences between groups in these variables at baseline.	Appropriate methods to control for measured confounders include stratification, regression, matching, standardization, and inverse probability weighting. They may control for individual variables or for the estimated propensity score. Each method depends on the assumption that there is no unmeasured or residual confounding.
1.6 Did the authors control for any post-intervention variables that could have been affected by the intervention?	NA/Y/PY/PN/N/NI Have the authors controlled for variables that are measured after the intervention is received? Are these variables likely to be affected by intervention or affect the outcome of the intervention?	Controlling for post-intervention variables that are affected by intervention is not appropriate. Controlling for mediating variables estimates the direct effect of intervention and may introduce bias. Controlling for common effects of intervention and outcome introduces bias.
1.7 Did the authors use an appropriate analysis method that adjusted for all the important confounding domains and for time-varying confounding?	NA/Y/PY/PN/N/NI Was there adjustment for differences between patients in adhering to ERT or any HSCT pre-treatment regimens?	Adjustment for time-varying confounding is necessary to estimate the effect of starting and adhering to intervention, in both randomized trials and NRSI. Appropriate methods include those based on inverse probability weighting. Standard regression models that include time-updated confounders may be problematic if time-varying confounding is present.
1.8 If Y/PY to 1.7: Were confounding domains that were adjusted for measured validly and reliably by the variables available in this study?	NA/Y/PY/PN/N/NI	

Question	Possible answer/reason for not using question	Guideline criteria
Risk of bias judgement	Low/ Moderate/ Serious/ Critical/ NI **Low**: no confounding expected **Moderate**: (i) Confounding expected, all known important confounding domains appropriately measured and controlled for; **and** (ii) Reliability and validity of measurement of important domains were sufficient, such that we do not expect serious residual confounding. **Serious**: (i) At least one known important domain was not appropriately measured, or not controlled for; **or** (ii) Reliability or validity of measurement of an important domain was low enough that we expect serious residual confounding. **Critical**: (i) Confounding inherently not controllable, **or** (ii) The use of negative controls strongly suggests unmeasured confounding.	This question should be answered after all studies have been considered, so a judgement can be made on how different domains could influence the estimate of the outcome, e.g. is there likely to be an impact of even a single uncontrolled domain on the outcome?
BIAS IN PARTICIPANT SELECTION		
2.1 Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention?	NA/Y/PY/PN/N/NI For example, were only participants with certain response or quality of life included in the analysis? Or, has the study started partway after the intervention was given? Consider that this is the case in retrospective studies, in which case, are participants likely to have been included in the study because they had certain characteristics predicting a positive outcome?	This domain is concerned only with selection into the study based on participant characteristics observed after the start of intervention. Selection based on characteristics observed before the start of intervention can be addressed by controlling for imbalances between experimental intervention and comparator groups in baseline characteristics that are prognostic for the outcome (baseline confounding). Is selection into the study based on intervention and/or outcome? I.e. were patients with only HSCT or only ERT recruited? Were patients with intervention done by a specific age recruited (e.g. if early is 5-7 years, were patients who had it at 5 years recruited ahead of those who had it at 7 years?)
2.2 If Y/PY to 2.1: Were the post-intervention variables that influenced selection likely to be associated with intervention? and, 2.3 If Y/PY to 2.2: Were the post-intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome?	NA/Y/PY/PN/N/NI Questions combined due to high similarity.	It is unlikely that participants would be included in the study on basis of anything else than having received a specific intervention (ERT, HSCT). However, if selection was also dependent on other inclusion criteria, were these associated with the intervention and if so, were these able to influence the outcome?

Question	Possible answer/reason for not using question	Guideline criteria
2.4 Do start of follow-up and start of intervention coincide for most participants?	NA/Y/PY/PN/N/NI	If participants are not followed from the start of the intervention then a period of follow up has been excluded, and individuals who experienced the outcome soon after intervention will be missing from analyses. This problem may occur when prevalent, rather than new (incident), users of the intervention are included in analyses.
2.5 If Y/PY to 2.2 and 2.3, or N/PN to 2.4: Were adjustment techniques used that are likely to correct for the presence of selection biases?	NA/Y/PY/PN/N/NI	It is in principle possible to correct for selection biases, for example by using inverse probability weights to create a pseudo-population in which the selection bias has been removed, or by modelling the distributions of the missing participants or follow up times and outcome events and including them using missing data methodology. However such methods are rarely used and the answer to this question will usually be “No”.
Risk of bias judgement	Low/ Moderate/ Serious/ Critical/ NI **Low**: (i) All participants who would have been eligible for the target trial were included in the study; and (ii) For each participant, start of follow up and start of intervention coincided. **Moderate**: (i) Selection into the study may have been related to intervention and outcome but appropriate methods to adjust for selection bias used or (ii) Start of follow up and start of intervention do not coincide for all participants; and either the proportion of participants for which this was the case was low or appropriate methods were used to account for this or it can be said with confidence the effect of intervention remains constant over time. **Serious**: (i) Selection into the study was related (but not very strongly) to intervention and outcome, and could not be adjusted for in analyses; or (ii) Start of follow up and start of intervention do not coincide and a potentially important amount of follow-up time is missing from analyses and the rate ratio is not constant over time **Critical**: (i) Selection into the study was very strongly related to intervention and outcome; and could not be adjusted for in analyses or (ii) A substantial amount of follow-up time is likely to be missing from analyses and the rate ratio is not constant over time.	
BIAS IN THE CLASSIFICATION OF INTERVENTIONS		
3.1 Were intervention groups clearly defined?	NA/Y/PY/PN/N/NI Was there a narrow range for "early" or "late" HSCT/ERT or were the definitions not stated? Could the groups of "early" and "late" overlap?	A pre-requisite for an appropriate comparison of interventions is that the interventions are well defined. Ambiguity in the definition may lead to bias in the classification of participants. For individual-level interventions, criteria for considering individuals to have received each intervention should be clear and explicit, covering issues such as type, setting, dose, frequency, intensity and/or timing of intervention.
Risk of bias judgement	Low/ Moderate/ Serious/ Critical/ NI **Low**: Intervention status is well defined and the definition is based	

Question	Possible answer/reason for not using question	Guideline criteria
	solely on information collected at the time of intervention. **Moderate**: Intervention status is well defined but some aspects of the assignments of intervention status were determined retrospectively. **Serious**: (i) Intervention status is not well defined; or (ii) Major aspects of the assignments of intervention status were determined in a way that could have been affected by knowledge of the outcome. **Critical**: (Unusual) An extremely high amount of misclassification of intervention status, e.g. because of unusually strong recall biases.	
BIAS DUE TO DEVIATIONS FROM INTENDED INTERVENTIONS		
4.1. Were there deviations from the intended intervention beyond what would be expected in usual practice?	NA/Y/PY/PN/N/NI	Deviations that happen in usual practice following the intervention (for example, cessation of a drug intervention because of acute toxicity) are part of the intended intervention and therefore do not lead to bias in the effect of assignment to intervention. Deviations may arise due to expectations of a difference between intervention and comparator (for example because participants feel unlucky to have been assigned to the comparator group and therefore seek the active intervention, or components of it, or other interventions). Such deviations are not part of usual practice, so may lead to biased effect estimates. However these are not expected in observational studies of individuals in routine care.
4.6. If N/PN to 4.3, 4.4 or 4.5: Was an appropriate analysis used to estimate the effect of starting and adhering to the intervention?	NA/Y/PY/PN/N/NI	It is possible to conduct an analysis that corrects for some types of deviation from the intended intervention. Examples of appropriate analysis strategies include inverse probability weighting or instrumental variable estimation. It is possible that a paper reports such an analysis without reporting information on the deviations from intended intervention, but it would be hard to judge such an analysis to be appropriate in the absence of such information. Specialist advice may be needed to assess studies that used these approaches. If everyone in one group received a co-intervention, adjustments cannot be made to overcome this.
Risk of bias judgement	Low/ Moderate/ Serious/ Critical/ NI **Low**: (i) Any deviations from intended intervention reflected usual practice; or (ii) Any deviations from usual practice were unlikely to impact on the outcome. **Moderate**: There were deviations from usual practice, but their impact on the outcome is expected to be slight. **Serious**: There were deviations from usual practice that were unbalanced between the intervention groups and likely to have affected the outcome.	

Question	Possible answer/reason for not using question	Guideline criteria
	Critical: There were substantial deviations from usual practice that were unbalanced between the intervention groups and likely to have affected the outcome.	
BIAS DUE TO MISSING DATA		
5.1 Were outcome data available for all, or nearly all, participants?	NA/Y/PY/PN/N/NI	"Nearly all" should be interpreted as "enough to be confident of the findings", and a suitable proportion depends on the context. In some situations, availability of data from 95% (or possibly 90%) of the participants may be sufficient, providing that events of interest are reasonably common in both intervention groups. One aspect of this is that review authors would ideally try and locate an analysis plan for the study.
5.2 Were participants excluded due to missing data on intervention status?	NA/Y/PY/PN/N/NI In e.g. chart review, participants with incomplete data on aspects of intervention received might be removed from analysis.	Missing intervention status may be a problem. This requires that the intended study sample is clear, which it may not be in practice.
5.3 Were participants excluded due to missing data on other variables needed for the analysis?	NA/Y/PY/PN/N/NI	This question relates particularly to participants excluded from the analysis because of missing information on confounders that were controlled for in the analysis.
5.4 If PN/N to 5.1, or Y/PY to 5.2 or 5.3: Are the proportion of participants and reasons for missing data similar across interventions?	NA/Y/PY/PN/N/NI	This aims to elicit whether either (i) differential proportion of missing observations or (ii) differences in reasons for missing observations could substantially impact on our ability to answer the question being addressed. "Similar" includes some minor degree of discrepancy across intervention groups as expected by chance.
5.5 If PN/N to 5.1, or Y/PY to 5.2 or 5.3: Is there evidence that results were robust to the presence of missing data?	NA/Y/PY/PN/N/NI	Evidence for robustness may come from how missing data were handled in the analysis and whether sensitivity analyses were performed by the investigators, or occasionally from additional analyses performed by the systematic reviewers. It is important to assess whether assumptions employed in analyses are clear and plausible. Both content knowledge and statistical expertise will often be required for this. For instance, use of a statistical method such as multiple imputation does not guarantee an appropriate answer. Review authors should seek naïve (complete-case) analyses for comparison, and clear differences between complete-case and multiple imputation-based findings should lead to careful assessment of the validity of the methods used.
Risk of bias judgement	Low/ Moderate/ Serious/ Critical/ NI **Low**: (i) Data were reasonably complete; or (ii) Proportions of and reasons for missing participants were similar across intervention groups; or (iii) The analysis addressed missing data and is likely to have removed any risk of bias. **Moderate**: (i) Proportions of and reasons for missing participants differ slightly across intervention groups; and (ii) The analysis is unlikely to have removed the risk of bias arising from the missing	

Question	Possible answer/reason for not using question	Guideline criteria
	data. **Serious**: (i) Proportions of missing participants or reasons for missingness differ substantially across interventions and (ii) The analysis is unlikely to have removed the risk of bias arising from the missing data or missing data were addressed inappropriately in the analysis or the nature of the missing data means that the risk of bias cannot be removed through appropriate analysis. **Critical**: (unusual) There were critical difference between interventions in participants with missing data and missing data were not (or could not) be addressed through appropriate analysis	
BIAS IN MEASUREMENT OF OUTCOMES		
6.1 Could the outcome measures have been influenced by knowledge of the intervention received?	NA/Y/PY/PN/N/NI I.e. would it matter if the study wasn't "blinded"? Question to be answered for each relevant reported outcome separately.	Some outcome measures involve negligible assessor judgment, e.g. all-cause mortality or non-repeatable automated laboratory assessments. Risk of bias due to measurement of these outcomes would be expected to be low.
6.2 Were outcome assessors aware of the intervention received by study participants?	NA/Y/PY/PN/N/NI I.e. was the person collecting the outcome data "blind" to the intervention? Question to be answered for each relevant reported outcome separately	If outcome assessors were blinded to intervention status, the answer to this question would be 'No'. In other situations, outcome assessors may be unaware of the interventions being received by participants despite there being no active blinding by the study investigators; the answer this question would then also be 'No'. In studies where participants report their outcomes themselves, for example in a questionnaire, the outcome assessor is the study participant. In an observational study, the answer to this question will usually be 'Yes' when the participants report their outcomes themselves.
6.3 Were the methods of outcome assessment comparable across intervention groups?	NA/Y/PY/PN/N/NI This is likely to be similar, but might actually differ if some patients were treated much later than others. It is mostly expected to be different for studies where a historical control is used or in before-and-after studies	Comparable assessment methods (i.e. data collection) would involve the same outcome detection methods and thresholds, same time point, same definition, and same measurements.
6.4 Were any systematic errors in measurement of the outcome related to intervention received?	NA/Y/PY/PN/N/NI Unless the outcomes were measure with different methods, this is highly unlikely.	This question refers to differential misclassification of outcomes. Systematic errors in measuring the outcome, if present, could cause bias if they are related to intervention or to a confounder of the intervention-outcome relationship. This will usually be due either to outcome assessors being aware of the intervention received or to non-comparability of outcome assessment methods, but there are examples of differential misclassification arising despite these controls being in place.
Risk of bias judgement	Low/ Moderate/ Serious/ Critical/ NI **Low**: The methods of outcome assessment were comparable across intervention groups and either the outcome measure was unlikely to be influenced by lack of blinding or outcome assessors were unaware of intervention received by study participants and error in	

Question	Possible answer/reason for not using question	Guideline criteria
	outcome measurement is unrelated to intervention status. **Moderate**: (i) The methods of outcome assessment were comparable across intervention groups; and (ii) The outcome measure is only minimally influenced by knowledge of the intervention received by study participants; and (iii) Any error in measuring the outcome is only minimally related to intervention status. **Serious**: (i) The methods of outcome assessment were not comparable across intervention groups; or (ii) The outcome measure was subjective and the outcome assessed by assessors aware of the intervention received by study participants; or (iii) Error in measuring the outcome was related to intervention status. **Critical**: The methods of outcome assessment were so different that they cannot reasonably be compared across intervention groups.	
BIAS IN SELECTION OF THE REPORTED RESULT		
Is the reported effect estimate likely to be selected, on the basis of the results, from... 7.1. ... multiple outcome measurements within the outcome domain?	NA/Y/PY/PN/N/NI E.g. if the number of people with specific threshold of outcome is reported, the threshold can be changed. This should be easy to spot if the threshold is different than what is usually used in other studies.	For a specified outcome domain, it is possible to generate multiple effect estimates for different measurements. If multiple measurements were made, but only one or a subset is reported, there is a risk of selective reporting on the basis of results.
.2 ... multiple analyses of the intervention-outcome relationship?	NA/Y/PY/PN/N/NI	Because of the limitations of using data from non-randomized studies for analyses of effectiveness (need to control confounding, substantial missing data, etc), analysts may implement different analytic methods to address these limitations. Examples include unadjusted and adjusted models; use of final value vs change from baseline vs analysis of covariance; different transformations of variables; a continuously scaled outcome converted to categorical data with different cut-points; different sets of covariates used for adjustment; and different analytic strategies for dealing with missing data. Application of such methods generates multiple estimates of the effect of the intervention versus the comparator on the outcome. If the analyst does not pre-specify the methods to be applied, and multiple estimates are generated but only one or a subset is reported, there is a risk of selective reporting on the basis of results.
7.3 ... different subgroups?	NA/Y/PY/PN/N/NI	Particularly with large cohorts often available from routine data sources, it is possible to generate multiple effect estimates for different subgroups or simply to omit varying proportions of the original cohort. If multiple estimates are generated but only one or a subset is reported, there is a risk of selective reporting on the basis of results.
Risk of bias judgement	Low/ Moderate/ Serious/ Critical/ NI **Low**: There is clear evidence (usually through examination of a pre-registered protocol or statistical analysis plan) that all reported	

Question	Possible answer/reason for not using question	Guideline criteria
	results correspond to all intended outcomes, analyses and sub-cohorts. **Moderate**: (i) The outcome measurements and analyses are consistent with an a priori plan; or are clearly defined and both internally and externally consistent; and (ii) There is no indication of selection of the reported analysis from among multiple analyses; and (iii) There is no indication of selection of the cohort or subgroups for analysis and reporting on the basis of the results. **Serious**: (i) Outcomes are defined in different ways in the methods and results sections, or in different publications of the study; or (ii) There is a high risk of selective reporting from among multiple analyses; or (iii) The cohort or subgroup is selected from a larger study for analysis and appears to be reported on the basis of the results. **Critical**: (i) There is evidence or strong suspicion of selective reporting of results; and (ii) The unreported results are likely to be substantially different from the reported results.	
OVERALL BIAS	Low/ Moderate/ Serious/ Critical/ NI Low: The study is judged to be at low risk of bias for all domains. Moderate: The study is judged to be at low or moderate risk of bias for all domains. Serious: The study is judged to be at serious risk of bias in at least one domain, but not at critical risk of bias in any domain. Critical: The study is judged to be at critical risk of bias in at least one domain.	

Table 20. Quality assessment of studies included for question 2

Question	Wyffels 2017	Eisengart 2018	Megens 2014	Wadhwa 2018	Aldenhoven 2015a	Laraway 2016	Javed 2018	Pal 2015	Poe 2014	Rodgers 2017	Kunin-Batson 2016	Langereis 2016	Aldenhoven 2015b
BIAS DUE TO CONFOUNDING													
1.1 Is there potential for confounding of the effect of intervention in this study?	Y	Y	Y	Y	Y	Y	Y	Y	Y	PY	PY	Y	PY
1.2 If Y/PY to 1.1: Was the analysis based on splitting participants' follow up time according to intervention received?	Y	Y	NA	NA	NA	NA	NA	N	NA	NA	NA	NA	NA

Question	Wyffels 2017	Eisengart 2018	Megens 2014	Wadhwa 2018	Aldenhoven 2015a	Laraway 2016	Javed 2018	Pal 2015	Poe 2014	Rodgers 2017	Kunin-Batson 2016	Langereis 2016	Aldenhoven 2015b
1.3 Were intervention discontinuations or switches likely to be related to factors that are prognostic for the outcome?	NI	PY	NA	NA	NA	PY	NA	NI	NA	NA	NA	NA	NA
1.4 Did the authors use an appropriate analysis method that controlled for all the important confounding domains?	N	N	N	N	PY	N	N	Y	Y	N	PY	N	PY
1.6 Did the authors control for any post-intervention variables that could have been affected by the intervention?	N	N	N	N	N	N	N	PY	N	N	N	N	N
1.7 Did the authors use an appropriate analysis method that adjusted for all the important confounding domains and for time-varying confounding?	N	N	NA	N	PN	N	N	N	N	PY	NA	N	PY
1.8 If Y/PY to 1.7: Were confounding domains that were adjusted for measured validly and reliably by the variables available in this study?	NA	NA	NA	NA	NA	NA	NA	NA	NA	PY	NA	NA	PY
Risk of bias judgement	Serious	Serious	Critical	Critical	Serious	Critical	Critical	Serious	Moderate	Moderate	Moderate	Serious	Low
BIAS IN PARTICIPANT SELECTION													
2.1 Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention?	Y	N	Y	Y	Y	PN	Y	Y	N	N	Y	Y	N

Question	Wyffels 2017	Eisengart 2018	Megens 2014	Wadhwa 2018	Aldenhoven 2015a	Laraway 2016	Javed 2018	Pal 2015	Poe 2014	Rodgers 2017	Kunin-Batson 2016	Langereis 2016	Aldenhoven 2015b
2.2 If Y/PY to 2.1: Were the post-intervention variables that influenced selection likely to be associated with intervention? **and,** **2.3 If Y/PY to 2.2: Were the post-intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome?**	2.2 PY 2.3 PY	NA	2.2 PY	2.2 Y 2.3 Y	PY	NA	PY	PY	NA	NA	Y	Y	NA
2.4 Do start of follow-up and start of intervention coincide for most participants?	PY	PY	PN	Y	PY	PY	PN	PY	Y	PY	PN	PY	PY
2.5 If Y/PY to 2.2 and 2.3, or N/PN to 2.4: Were adjustment techniques used that are likely to correct for the presence of selection biases?	N	NA	N	NA	N	NA	N	N	NA	NA	N	N	NA
Risk of bias judgement	Critical	Low	Serious	Serious	Serious	Serious	Serious	Serious	Low	Low	Serious	Serious	Low
BIAS IN THE CLASSIFICATION OF INTERVENTIONS													
3.1 Were intervention groups clearly defined?	PN	N	N	Y	PY	PY	PN	N	PN	Y	N	N	PN
Risk of bias judgement	Serious	Moderate	Serious	Low	Serious	Moderate	Moderate	Serious	Moderate	Low	Serious	Serious	Serious
BIAS DUE TO DEVIATIONS FROM INTENDED INTERVENTIONS													
4.1. Were there deviations from the intended intervention beyond what would be expected in usual practice?	PN	N	N	N	N	N	N	PN	N	N	N	N	PN
4.6. If N/PN to 4.3, 4.4 or 4.5:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Question	Wyffels 2017	Eisengart 2018	Megens 2014	Wadhwa 2018	Aldenhoven 2015a	Laraway 2016	Javed 2018	Pal 2015	Poe 2014	Rodgers 2017	Kunin-Batson 2016	Langereis 2016	Aldenhoven 2015b
Was an appropriate analysis used to estimate the effect of starting and adhering to the intervention?													
Risk of bias judgement	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low
BIAS DUE TO MISSING DATA													
5.1 Were outcome data available for all, or nearly all, participants?	NI	Y	Y	Y	NI	Y	N	Y	PY	Y	N	PY	Y
5.2 Were participants excluded due to missing data on intervention status?	NI	NI	NI	PN	NI	PN	NI	PN	PN	NI	Y	PY	PN
5.3 Were participants excluded due to missing data on other variables needed for the analysis?	Y	Y	PY	PN	NI	PN	NI	PN	PN	PN	Y	Y	PN
5.4 If PN/N to 5.1, or Y/PY to 5.2 or 5.3: Are the proportion of participants and reasons for missing data similar across interventions?	NI	PN	NA	NA	NA	PN	NI	NA	NA	NA	NA	NA	NA
5.5 If PN/N to 5.1, or Y/PY to 5.2 or 5.3: Is there evidence that results were robust to the presence of missing data?	PN	PN	PN	NA	N	NA	PN	NA	NA	NA	N	N	NA
Risk of bias judgement	Serious	Moderate	Moderate	Low	Serious	Low	Serious	Low	Low	Low	Serious	Low	Low
BIAS IN MEASUREMENT OF OUTCOMES													
6.1 Could the outcome measures have been influenced by knowledge of the intervention received?	N	PY	N	Survival: N Disease-related mortality: PY	Neurological development: PN Secondary	Mitral/aortic valve function and corneal clouding:	Corneal clouding: PY Visual acuity: PN	N	N	N	PN	PY	PN

Question	Wyffels 2017	Eisengart 2018	Megens 2014	Wadhwa 2018	Aldenhoven 2015a	Laraway 2016	Javed 2018	Pal 2015	Poe 2014	Rodgers 2017	Kunin-Batson 2016	Langereis 2016	Aldenhoven 2015b
					outcomes: PY	PY Visual acuity: PN							
6.2 Were outcome assessors aware of the intervention received by study participants?	NI	PY	PY	PY	PY	Y	PY	PY	Y	PY	PY	N	PY
6.3 Were the methods of outcome assessment comparable across intervention groups?	PY	Survival: Y CNS outcomes: PY	Y	PY	PN	PY	PY	PY	Y	NA	N	NA	NA
6.4 Were any systematic errors in measurement of the outcome related to intervention received?	N	PN	NA	NA	PN	N	N	N	N	NA	PN	PN	PN
Risk of bias judgement	Low	Low	Low	Moderate	Serious	Serious	Serious	Low	Low	Low	Moderate	Low	Low
BIAS IN SELECTION OF THE REPORTED RESULT													
Is the reported effect estimate likely to be selected, on the basis of the results, from... **7.1. ... multiple outcome measurements within the outcome domain?**	N	PN	N	N	Neurological development: PY Secondary outcomes: PN	PN	N	N	PN	PN	N	PN	N
.2 ... multiple analyses of the intervention-outcome relationship?	PN	PN	N	N	PN	N	N	N	N	PN	Y	PN	N
7.3 ... different subgroups?	N	PY	N	N	PN	N	NA	PY	N	N	N	N	N
Risk of bias judgement	Moderate	Serious	Moderate	Moderate	Serious	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate
OVERALL BIAS	**Critical**	**Serious**	**Critical**	**Critical**	**Serious**	**Critical**	**Critical**	**Serious**	**Moderate**	**Moderate**	**Serious**	**Serious**	**Moderate**

Appendix 6 – 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 21.

Table 21. UK NSC reporting checklist for evidence summaries

	Section	Item	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.	6–9
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 Objectives – What are the questions the current evidence summary intends to answer? – statement of the key questions for the current evidence summary, criteria they address, and number of studies included per question, description of the overall results of the literature search.	10–18

		Method – briefly outline the rapid review methods used.	
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*.	18–21
2.3	Appraisal for quality/risk of bias tool	Details of tool/checklist used to assess quality, e.g. QUADAS 2, CASP, SIGN, AMSTAR.	22
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.	22
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.	51–53
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.	54–56
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.). 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.	**Study level reporting:** Q1: 29 Q2: 38–40 **Quality assessment:** Q1: 25–28 Q2: 35–37

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.	Q1: 23–25 Q2: 33-34
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.	Q1: 30 Q2: 42–43
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'?	Q1: 31–32 Q2: 44–45
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?	46–49
6.2	Limitations	Discuss limitations of the available evidence and of the review methodology if relevant.	49–50

References

1. Wraith JE. The mucopolysaccharidoses: a clinical review and guide to management. Arch Dis Child 1995;72:263-7.
2. Kingma SD, Langereis EJ, de Klerk CM, et al. An algorithm to predict phenotypic severity in mucopolysaccharidosis type I in the first month of life. Orphanet J Rare Dis 2013;8:99.
3. Beck M, Arn P, Giugliani R, et al. The natural history of MPS I: global perspectives from the MPS I Registry. Genet Med 2014;16:759-65.
4. Clarke LA. Mucopolysaccharidosis Type I [Updated 2016 Feb 11]. In: Adam MP, Ardinger HH, Pagon RA, et al., editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2019, 2002.
5. UK NSC. Newborn Screening for Mucopolysaccharidosis I: External Review against Programme Appraisal Criteria for the UK National Screening Committee, 2015.
6. D'Aco K, Underhill L, Rangachari L, et al. Diagnosis and treatment trends in mucopolysaccharidosis I: findings from the MPS I Registry. Eur J Pediatr 2012;171:911-9.
7. Thomas JA, Beck M, Clarke JT, et al. Childhood onset of Scheie syndrome, the attenuated form of mucopolysaccharidosis I. J Inherit Metab Dis 2010;33:421-7.
8. Moore D, Connock MJ, Wraith E, et al. The prevalence of and survival in Mucopolysaccharidosis I: Hurler, Hurler-Scheie and Scheie syndromes in the UK. Orphanet J Rare Dis 2008;3:24.
9. Orphanet. Mucopolysaccharidosis type 1.
10. Genetic and Rare Diseases Information Center. Mucopolysaccharidosis type I.
11. Vijay S, Wraith JE. Clinical presentation and follow-up of patients with the attenuated phenotype of mucopolysaccharidosis type I. Acta Paediatr 2005;94:872-7.
12. Cleary M, Wraith J. The presenting features of mucopolysaccharidosis type IH (Hurler syndrome). Acta Paediatrica 1995;84:337-339.
13. Genzyme Corporation. Diagnosis and Testing, 2019.
14. Gabrielli O, Clarke LA, Bruni S, et al. Enzyme-replacement therapy in a 5-month-old boy with attenuated presymptomatic MPS I: 5-year follow-up. Pediatrics 2010;125:e183-7.
15. Wold SM, Derkay CS, Darrow DH, et al. Role of the pediatric otolaryngologist in diagnosis and management of children with mucopolysaccharidoses. International Journal of Pediatric Otorhinolaryngology 2010;74:27-31.
16. Wang RY, Bodamer OA, Watson MS, et al. Lysosomal storage diseases: Diagnostic confirmation and management of presymptomatic individuals. Genetics In Medicine 2011;13:457.
17. Pastores G, A Meere P. Musculoskeletal complications associated with lysosomal storage disorders: Gaucher disease and Hurler-Scheie syndrome (mucopolysaccharidosis type I), 2005.
18. Clarke L, Atherton A, Burton B, et al. Mucopolysaccharidosis Type I Newborn Screening: Best Practices for Diagnosis and Management, 2016.
19. Kemper A. Newborn Screening for MPS I: Final Report from the Condition Review Workgroup, 2015.
20. Parini R, Deodato F, Di Rocco M, et al. Open issues in Mucopolysaccharidosis type I-Hurler. Orphanet Journal of Rare Diseases 2017;12:112.
21. Viskochil D. Mucopolysaccharidosis Type I (MPS 1), 2019.
22. Donati MA, Pasquini E, Spada M, et al. Newborn screening in mucopolysaccharidoses. Italian Journal of Pediatrics 2018;44:126.

23. Tanpaiboon P. Lysosomal storage disorders. 2014;4:217-229.
24. Advisory Committee on Heritable Disorders in Newborns and Children. 2017 Report to Congress, 2017.
25. Hopkins PV, Campbell C, Klug T, et al. Lysosomal storage disorder screening implementation: Findings from the first six months of full population pilot testing in Missouri. Journal of Pediatrics 2015;166:172-177.
26. Burton BK, Charrow J, Hoganson GE, et al. Newborn Screening for Lysosomal Storage Disorders in Illinois: The Initial 15-Month Experience. J Pediatr 2017;190:130-135.
27. Scott CR, Elliott S, Buroker N, et al. Identification of infants at risk for developing Fabry, Pompe, or mucopolysaccharidosis-I from newborn blood spots by tandem mass spectrometry. J Pediatr 2013;163:498-503.
28. Lin SP, Lin HY, Wang TJ, et al. A pilot newborn screening program for Mucopolysaccharidosis type I in Taiwan. Orphanet J Rare Dis 2013;8:147.
29. Bravo H, Neto EC, Schulte J, et al. Investigation of newborns with abnormal results in a newborn screening program for four lysosomal storage diseases in Brazil. Mol Genet Metab Rep 2017;12:92-97.
30. Paciotti S, Persichetti E, Pagliardini S, et al. First pilot newborn screening for four lysosomal storage diseases in an Italian region: identification and analysis of a putative causative mutation in the GBA gene. Clin Chim Acta 2012;413:1827-31.
31. Burlina AB, Polo G, Salviati L, et al. Newborn screening for lysosomal storage disorders by tandem mass spectrometry in North East Italy. Journal of Inherited Metabolic Disease 2018;41:209-219.
32. Navarrete-Martinez JI, Limon-Rojas AE, Gaytan-Garcia MJ, et al. Newborn screening for six lysosomal storage disorders in a cohort of Mexican patients: Three-year findings from a screening program in a closed Mexican health system. Mol Genet Metab 2017;121:16-21.
33. Hoffmann GF, Lindner M, Loeber JG. 50 years of newborn screening. J Inherit Metab Dis 2014;37:163-4.
34. National Health Service. Newborn blood spot test, 2018.
35. Minter Baerg MM, Stoway SD, Hart J, et al. Precision newborn screening for lysosomal disorders. Genetics in Medicine 2018;20:847-854.
36. Chuang CK, Lin HY, Wang TJ, et al. Status of newborn screening and follow up investigations for Mucopolysaccharidoses i and II in Taiwan. Orphanet Journal of Rare Diseases 2018;13 (1) (no pagination).
37. Aldenhoven M, Jones SA, Bonney D, et al. Hematopoietic Cell Transplantation for Mucopolysaccharidosis Patients Is Safe and Effective: Results after Implementation of International Guidelines. Biology of Blood and Marrow Transplantation 2015;21:1106-1109.
38. Aldenhoven M, Wynn RF, Orchard PJ, et al. Long-term outcome of Hurler syndrome patients after hematopoietic cell transplantation: An international multicenter study. Blood 2015;125:2164-2172.
39. Wyffels ML, Orchard PJ, Shanley RM, et al. The Frequency of Carpal Tunnel Syndrome in Hurler Syndrome After Peritransplant Enzyme Replacement Therapy: A Retrospective Comparison. Journal of Hand Surgery 2017;42:573.e1-573.e8.
40. Rodgers NJ, Kaizer AM, Miller WP, et al. Mortality after hematopoietic stem cell transplantation for severe mucopolysaccharidosis type I: the 30-year University of Minnesota experience. Journal of Inherited Metabolic Disease 2017;40:271-280.
41. Wadhwa A, Chen Y, Holmqvist A, et al. Late Mortality after Allogeneic Blood or Marrow Transplantation for Inborn Errors of Metabolism: A Report from the Blood or Marrow

Transplant Survivor Study-2 (BMTSS-2). Biology of Blood and Marrow Transplantation 2019;25:328-334.
42. Poe MD, Chagnon SL, Escolar ML. Early treatment is associated with improved cognition in hurler syndrome. Annals of Neurology 2014;76:747-753.
43. Javed A, Aslam T, Jones SA, et al. The effect of haemopoietic stem cell transplantation on the ocular phenotype in mucopolysaccharidosis type I (Hurler). Acta Ophthalmologica 2018;96:494-498.
44. Laraway S, Mercer J, Jameson E, et al. Outcomes of Long-Term Treatment with Laronidase in Patients with Mucopolysaccharidosis Type I. Journal of Pediatrics 2016;178:219-226.e1.
45. Megens JHAM, De Wit M, Van Hasselt PM, et al. Perioperative complications in patients diagnosed with mucopolysaccharidosis and the impact of enzyme replacement therapy followed by hematopoietic stem cell transplantation at early age. Paediatric Anaesthesia 2014;24:521-527.
46. Pal AR, Langereis EJ, Saif MA, et al. Sleep disordered breathing in mucopolysaccharidosis I: A multivariate analysis of patient, therapeutic and metabolic correlators modifying long term clinical outcome. Orphanet Journal of Rare Diseases 2015;10 (1) (no pagination).
47. Eisengart JB, Rudser KD, Xue Y, et al. Long-term outcomes of systemic therapies for Hurler syndrome: an international multicenter comparison. Genetics in Medicine 2018;20:1423-1429.
48. Kunin-Batson AS, Shapiro EG, Rudser KD, et al. Long-Term Cognitive and Functional Outcomes in Children with Mucopolysaccharidosis (MPS)-IH (Hurler Syndrome) Treated with Hematopoietic Cell Transplantation. Jimd Reports 2016;29:95-102.
49. Langereis EJ, Den Os MM, Breen C, et al. Progression of hip dysplasia in mucopolysaccharidosis type ihurler after successful hematopoietic stem cell transplantation. Journal of Bone and Joint Surgery - American Volume 2016;98:386-395.
50. Lutgendorf MA, Stoll KA. Why 99% may not be as good as you think it is: limitations of screening for rare diseases. J Matern Fetal Neonatal Med 2016;29:1187-9.
51. Lalkhen AG MA. Clinical tests: sensitivity and specificity. Continuing Education and Anesthesia Critical Care & Pain 2008;8:221-223.