



Newborn screening for metachromatic leukodystrophy (MLD): A rapid evidence review

External review against programme appraisal criteria for the UK National Screening Committee (NSC), prepared by Kleijnen Systematic Reviews Ltd

Version: 14 (Final)

Authors: Westwood M E, Richmond Mynett Z, Chen J, Armstrong N, Noake C

Date: 28 November 2025

The UK NSC secretariat is hosted by the Department of Health and Social Care

Contents

Newborn screening for metachromatic leukodystrophy (MLD): A rapid evidence review	1
Contents	2
List of tables.....	4
About the UK National Screening Committee (UKNSC)	5
Plain English summary	6
Executive summary.....	8
Purpose of the review.....	8
Background	8
Focus of the review	9
Recommendation under review	10
Findings and gaps in the evidence of this review.....	10
Recommendations on screening	11
Limitations	12
Introduction and approach	13
Background	13
Current policy context and previous reviews.....	14
Objectives	16
Methods	18
Eligibility for inclusion in the review.....	19
Data extraction	23
Appraisal for quality/risk of bias tool	23
Methods of analysis/synthesis	23
Databases/sources searched	23
Horizon scanning searches	24
Additional searches	24
Update searches	25
Overview of included studies	26
Question level synthesis	28
Criteria 4 and 5 - Accuracy of the screening test	28
Summary of findings relevant to Criteria 4 and 5	40
Criterion 9 - Efficacy of treatment in the pre-symptomatic phase	41
Post-consultation addendum	46
Summary of Findings Relevant to Criterion 9	48
Criterion 14 - Cost effectiveness of NBS for MLD.....	62
Summary of findings relevant to Criterion 14	65
Review summary	66
Conclusions and implications for policy	66

Limitations	67
Appendix 1 - Search strategy	68
Electronic databases	68
Search terms	70
Update searches	80
Appendix 2 - Included and excluded studies	83
PRISMA flowchart.....	83
Appendix 3 - Appraisal for quality and risk of bias of individual studies	90
QUADAS-2 and QUADAS C assessments.....	90
ROBINS-I assessments	101
Appendix 4 - Screening algorithms evaluated by included studies	117
Germany	117
UK.....	118
US (Washington State)	119
Appendix 5 - Published clinical guidelines on the management of MLD	121
NICE guidance HST18: Atidarsagene autotemcel for treating metachromatic leukodystrophy	121
Newborn screening in MLD – European consensus-based recommendations on clinical management	121
Appendix 6 - UK NSC reporting checklist for evidence summaries	124
References	131

List of tables

Table 1: Key questions for the evidence summary and relationship to the UKNSC screening criteria.....	17
Table 2: Inclusion and exclusion criteria for the key questions.....	20
Table 3: Summary of the screening algorithms evaluated in included studies	35
Table 4: Summary of screening algorithm performance.....	37
Table 5: Details of clinically relevant variants identified in screen-positive study participants.....	38
Table 6: Summary of QUADAS-2 and QUADAS-C evaluations.....	39
Table 7: Details of included treatment studies	50
Table 8: Participant characteristics for treatment studies	52
Table 9: Treatment details	54
Table 10: Dichotomous outcomes from treatment studies	55
Table 11: Continuous outcomes from treatment studies	59
Table 12: Summary of ROBINS-I assessments	61
Table 13: Summary methodological quality assessment using then Drummond checklist	63
Table 14: Databases searched.....	68
Table 15: Search strategy for MEDLINE, MEDLINE In-Process, MEDLINE Daily, Epub Ahead of Print (Searched via Ovid).....	70
Table 16: Search strategy for Embase (Searched via Ovid)	71
Table 17: Search strategy for CINAHL (Searched via EBSCO)	72
Table 18: Search strategy for Orphanet (Searched via https://www.orpha.net).....	73
Table 19: Search strategy for Orphanet: Newborn Screening Bibliographical Knowledgebase (Searched via https://nbs.orphanet.app/?lang=en).....	73
Table 20: Search strategy for CDSR and CENTRAL (Searched via The Cochrane Library [Wiley])	73
Table 21: Search strategy for KSREvidence (searched via https://ksrevidence.com/).....	74
Table 22: Search strategy for NIH ClinicalTrials.gov (searched via http://www.clinicaltrials.gov/).....	74
Table 23: Search strategy for EU Clinical Trials Register (EUCTR) (Searched via https://www.clinicaltrialsregister.eu/ctr-search/search):.....	75
Table 24: Search strategy for World Health Organization International Clinical Trials Registry Platform (WHO ICTRP) (Searched via http://www.who.int/ictrp/en/).....	76
Table 25: Search strategy for ScanMedicine (Searched via https://scanmedicine.com/).....	77
Table 26: Search strategy for TRIP database (Searched via https://www.tripdatabase.com/).....	77
Table 27: Search strategy for International Guidelines Library (GIN) (searched via https://g-i-n.net/international-guidelines-library/).....	78
Table 28: Search strategy for National Institute for Health and Care Excellence (NICE) (Searched via https://www.nice.org.uk/)	78
Table 29: Search strategy for the International HTA Database (INAHTA) (searched via https://database.inahta.org/).....	79
Table 30: Search strategy for National Institute for Health and Care Research (NIHR) HTA (Searched via https://www.nihr.ac.uk/):	79
Table 31: Search strategy for ECRI Guidelines Trust (Searched via https://guidelines.ecri.org/).....	80
Table 32: Update Searches	80
Table 33: Search strategy for MEDLINE, MEDLINE In-Process, MEDLINE Daily, Epub Ahead of Print (Searched via Ovid).....	80
Table 34: Search strategy for Embase (Searched via Ovid)	81
Table 35: Search strategy for medRxiv (Searched via https://www.medrxiv.org/)	82
Table 36: Summary of publications included after review of full-text articles, and the question(s) each publication was identified as being relevant to	84
Table 37: Summary of the identification and inclusion or exclusion of publications identified by the 2023 evidence map	85
Table 38: Publications excluded after review of full-text articles	88
Table 39: Reporting checklist	124

About the UK National Screening Committee (UKNSC)

The UKNSC advises ministers and the National Health Service (NHS) in the four United Kingdom (UK) countries about all aspects of [population screening](#) and supports implementation of screening programmes.

Conditions are reviewed against [evidence review criteria](#) according to the UKNSC's [evidence review process](#).

Read a [complete list of UKNSC recommendations](#).

UK National Screening Committee, Southside, 39 Victoria Street, London, SW1H 0EU

www.gov.uk/uknsc

Blog: <https://nationalscreening.blog.gov.uk/>

For queries relating to this document, please contact: <https://view-health-screening-recommendations.service.gov.uk/helpdesk/>

© Crown copyright 2025

You may re-use this information (excluding logos) free of charge in any format or medium, under the terms of the Open Government Licence v3.0. To view this licence, visit OGL or email psi@nationalarchives.gsi.gov.uk. Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

Published January 2026

Plain English summary

When a new population screening programme is proposed in the United Kingdom (UK), it is assessed using the UK National Screening Committee (NSC) criteria for appraising its viability, effectiveness and appropriateness. The overall goal of population screening programmes is to provide early treatment or intervention to someone identified as being at higher risk of a condition before they have symptoms. Ideally this should lead to better outcomes than if the person were to present later with symptoms. In the UK, the current newborn screening programme looks for nine rare but serious conditions. Screening for a 10th condition was recommended by the UK NSC in 2022 and is in the process of being implemented. Screening uses drops of blood, collected from an infant's heel onto a special card (also known as the 'heel prick test'). In the rare event that laboratory tests on this blood find an abnormal result, the child undergoes further testing to confirm whether or not they have one of the conditions screened for. If a child is then diagnosed with one of the conditions, they are referred for treatment.

Metachromatic leukodystrophy (MLD) is a rare, inherited condition that results in progressive nerve damage leading to muscle weakness, loss of co-ordination and mobility and loss of cognitive function, which worsens over time. MLD is usually classified according to age at which symptoms develop: late infantile (typically before 30 months of age), juvenile (typically between 3 and 16 years of age), and adult (typically after 16 years of age). The late infantile form is the most common and most severe form of MLD and usually results in death before the age of 5 years.

MLD is usually found after birth, once a baby shows symptoms, unless there is known history of MLD in the family (e.g. a previously diagnosed brother or sister).

In 2022, the National Institute for Health and Care Excellence (NICE) recommended a new treatment (Libmeldy®) for the late infantile and early juvenile forms of MLD, to be used before symptoms develop or when early symptoms are present. During the development of this recommendation, clinicians and patients highlighted the importance of early diagnosis and newborn screening for inherited disorders such as MLD.

Screening for MLD is not currently included in the UK newborn screening programme and has not previously been considered by the UKNSC.

There are a number of tests which could potentially be used to screen for MLD, and screening could also identify babies with multiple sulfatase deficiency (MSD), another ultra-rare inherited condition.

The aim of this 2025 evidence summary was to assess the published evidence relevant to newborn screening for MLD. It was commissioned by the UKNSC, following an initial assessment, which concluded that there was sufficient evidence to justify a more in-depth review of the evidence and that MLD should be added to the UKNSC's recommendation list, to be kept under regular review.

This evidence summary looked at evidence on the accuracy of tests that could be used to screen for MLD and the effectiveness of treatments in babies identified by screening. It also considered evidence about whether introduction of a screening programme for MLD represents value for money, in the context of available resource in the UK National Health Service (NHS).

The 2025 evidence summary has concluded that key areas of uncertainty remain over the best way to identify babies with MLD and whether identification of babies with MLD through newborn screening results in more effective treatment with better long-term outcomes. Experience from any implemented screening programmes (e.g. Norway) and ongoing pilot studies may provide evidence to reduce uncertainty in the future.

Executive summary

Purpose of the review

The overall aim of this project was to summarise the available evidence relevant to newborn screening for metachromatic leukodystrophy (MLD). This evidence summary was commissioned by the United Kingdom (UK) National Screening Committee (NSC), following completion of an evidence map on newborn screening for MLD in 2023.

Background

MLD, also known as Arylsulfatase A (ARSA) deficiency, is a rare neurodegenerative disease, in which deficiency in the ARSA enzyme leads to accumulation of sulfatides and consequent damage to the myelin sheath of neurons. MLD is a lysosomal storage disorder with autosomal recessive inheritance. MLD has three forms which are classified according to age at symptom onset: late infantile (typically presenting before 30 months of age), juvenile (typically presenting between 3 and 16 years of age), and adult (typically presenting after 16 years of age), with late infantile being the most common (50%-60% of cases) and most severe form. The incidence of MLD in the UK has been estimated at approximately 1:40,000 live births.

MLD is usually detected after birth and once symptoms have manifested, unless there is an awareness of family history/mutation status or previous development of MLD in a sibling.

Potential methods of screening for MLD utilise the measurement of sulfatide levels in urine or dried blood spot (DBS) samples, the quantification (by immunoassay) of ARSA protein abundance in DBS and the measurement of ARSA enzymatic activity in DBS samples using tandem mass spectrometry. Screening for MLD also has the potential to identify individuals with multiple sulfatase deficiency (MSD), another ultra-rare lysosomal storage disorder where affected patients also display high sulfatide levels and low ARSA enzymatic activity in the blood.

Atidarsagene autotemcel (ARSA-cel/OTL-200, developed by Orchard Therapeutics and branded as Libmeldy®) was recommended in 2022 by the National Institute for Health and Care Excellence (NICE), Highly Specialised Technology (HST) guidance 18 (HST18), as an option for treating MLD in presymptomatic children with late infantile or early juvenile MLD, and in children with early juvenile MLD who have early clinical signs or symptoms (who can still walk independently and who have no cognitive decline). Libmeldy® was approved by the European Medicines Agency (EMA) in 2020 and by the U.S. Food and Drug Administration (FDA) in March 2024 for the treatment of presymptomatic late infantile, presymptomatic early juvenile or early symptomatic early juvenile MLD.

Routine newborn screening for MLD is not currently recommended by the UKNSC and MLD is not included in the Recommended Uniform Screening Panel (RUSP) in the United States (US).

Screening for MLD has not previously been considered by the UKNSC. It was proposed as a potential newborn blood spot (NBS) population screening programme in the 2021 annual call for topics. The submission reasoned that, without screening, affected individuals are only identified before symptom onset when an older sibling is affected and that this limits the opportunity for treatment in individuals without affected siblings. In 2023, a preliminary evidence map was commissioned by the UKNSC to evaluate the volume and type of evidence related to newborn screening for MLD. The evidence map concluded that there was sufficient evidence to justify

commissioning an evidence summary and that MLD should be added to the UKNSC's recommendation list, to be kept under regular review.

Focus of the review

This evidence summary considered the evidence to inform four UKNSC criteria for a population screening programme. The criteria considered and the associated research questions were as follows:

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

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

Research question 1: What is the accuracy of single test and 2-tier NBS screening strategies for MLD, using DBS samples?

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.

Research question 2: Does early initiation of treatment following screening lead to improved outcomes for MLD compared to initiation of treatment following clinical presentation?

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

Research question 3: How have modelling studies and CEAs addressed NBS screening for MLD in the era of novel treatments?

In addition to summarising the available evidence to inform the above questions, this report includes:

- a horizon scanning exercise to identify any ongoing studies and recent developments in novel therapies for MLD
- information about any implemented NBS screening programmes for MLD that are relevant to the UK context
- a summary of clinical guidelines on the management of MLD that are relevant to the UK context

In order to maintain relevance to current clinical practice, this evidence summary considers research published since 2012.

Recommendation under review

Newborn screening for MLD has not previously been considered by the UKNSC.

Findings and gaps in the evidence of this review

Criterion 4 (There should be a simple, safe, precise and validated screening test)

The available evidence to inform research question 1 '*What is the accuracy of single test and 2-tier NBS screening strategies for MLD, using DBS samples?*' was sparse. All three publications included in this evidence summary reported early-stage studies which aimed to assess the feasibility of implementing NBS screening for MLD and all three studies were rated as having high risk of bias with respect to evaluating the accuracy of NBS screening algorithms for MLD.

It is important to note that no study included in this evidence summary reported either confirmatory genetic testing of screen negative DBS or any method (e.g. records review or surveillance) designed to identify cases of MLD that may have been missed by screening (false negative [FN]). Hence all reported or calculated estimates of the performance of NBS screening algorithms for MLD are uncertain and speculative in nature, since they assume that no cases of MLD were missed.

We did not identify any studies which reported experience from implemented screening programmes.

The limited evidence currently available indicates that criterion 4 is currently not met. This conclusion is based on the lack of data to inform estimates of the accuracy of evaluated screening algorithms and the high variability in estimates of PPV calculated from two studies, for the same 2-tier screening algorithm.

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

Findings from the small UK 'pre-pilot' study included in this evidence summary indicate that criterion 5 is currently not met. This conclusion is based on the incidental identification of a new case of late infantile MLD, during the validation phase of this study; DBS from this newborn had a C16:0- sulfatide level of 0.15 $\mu\text{mol/L}$, which is below the ≥ 0.17 $\mu\text{mol/L}$ cut-off used in the 2-tier algorithm evaluated by all three of the studies included in this evidence summary and which has been reported as the cut-off required to achieve 100% sensitivity.

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)

The limited evidence currently available indicates that criterion 9 is currently not met. There is some very weak, indirect evidence to indicate that the effects of gene therapy treatment (Libmeldy®) on gross motor function, relative to untreated patients, may be greater where patients receive treatment before symptoms develop. This evidence is derived from two published analyses comparing the outcomes of treated patients from one small study (NC T01560182), and from longer-term follow up of participants in NCT01560182 and a separate open label study (NCT03392987), to those of an untreated natural history cohort; these

analyses had substantial methodological limitations and were partially funded by Orchard Therapeutics (the manufacturer of Libmeldy®). The earlier analysis also formed the basis of the company's submission for NICE HST18. Even with the inclusion of indirect evidence, and the additional consideration of Fumagalli et al (2025), the available evidence base is small in volume and has methodological limitations. There is currently no direct evidence that identification of patients with MLD through screening or cascade testing results in improved outcomes relative to treatment following symptomatic presentation.

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

The available evidence to inform research question 3 'How have modelling studies and CEAs addressed NBS screening for MLD in the era of novel treatments?' is derived from a single publication, which reports an economic evaluation of MLD screening in the UK with substantial methodological limitations. The lead author of this publication and three additional study authors were employees of Orchard Therapeutics, the company that manufactures ARSA-cel (Libmeldy®), and two further authors received payment from Orchard Therapeutics for Markov model development.

Reported findings indicate that newborn screening can significantly increase the number of presymptomatic MLD patients diagnosed within the treatment window, allowing for earlier intervention with ARSA-cel (Libmeldy®), which is associated with substantial improvements in survival and quality of life (QoL). Sensitivity analyses tested variations in incidence rates, treatment eligibility, and discount rates, demonstrating that newborn screening remains cost-effective under most scenarios. However, there was a lack of justification in the choice of model parameters and ranges for sensitivity analyses, that could impact result reliability. Crucially, the reliance on clinical expert opinion for several parameters and the lack of transparency in the source of the parameters that were key drivers i.e. the treatment effect of ARSA-cel (Libmeldy®) means that the robustness of the findings is questionable. Overall, these findings provide the most comprehensive published economic evaluation of MLD screening in the UK. However the rapid review format has limitations and further work, either through an external quality assurance exercise of and changes to the existing model or the development of a de novo model, would be required to comment more fully on whether this criterion is met or not met (i.e. criterion 14 is 'uncertain').

Recommendations on screening

The current published evidence base alone is not adequate to support implementation of NBS screening for MLD.

Future publication of data from implemented screening programmes and ongoing pilot studies has the potential to provide evidence to inform criteria 4 and 5.

Further work is needed to adequately evaluate the performance of screening algorithms for MLD, in practice, and to establish the cut-off values appropriate for use in the UK population. Methodologically robust studies are needed to confirm the clinical effectiveness of available treatments for MLD and to test the hypothesis that treatment outcomes are improved where patients are treated before the onset of symptoms (i.e. through screening). Evidence about the

performance of screening algorithms and the efficacy of treatment is a pre-requisite to provide robust model inputs for CEAs.

Limitations

The paucity and poor quality of evidence, across all the criteria considered in this evidence summary, is a key limitation. Evidence generation is still at a relatively early stage and ongoing pilot studies and/or data collection from the first implemented screening programmes are likely to inform future evidence reviews.

This evidence summary employed standard systematic review methodology to ensure that the capture of relevant evidence was as complete as possible. In addition, to provide further context, we sought information about existing guidelines and any implemented NBS screening programmes for MLD. We also conducted a horizon scanning exercise to identify any ongoing studies of novel treatments for MLD.

The systematic review component of this evidence summary was limited by a restriction to full publications in the English language.

Introduction and approach

Background

Metachromatic leukodystrophy (MLD), also known as Arylsulfatase A (ARSA) deficiency, is a rare neurodegenerative disease, in which deficiency in the ARSA enzyme leads to the accumulation of sulfatides and consequent damage to the myelin sheath of neurons.^{1, 2} MLD is a lysosomal storage disorder with autosomal recessive inheritance.^{1, 2} The incidence of MLD in the United Kingdom (UK) has been estimated at approximately 1:40,000 live births.³ MLD has three forms which are classified according to age at symptom onset: late infantile (typically presenting before 30 months of age), juvenile (typically presenting between 3 and 16 years of age), and adult (typically presenting after 16 years of age).¹ The late infantile form is the most severe and most common form of MLD, comprising 50%-60% of cases. Rapid progression of the late infantile form of MLD usually results in death before the age of 5 years.^{2, 4}

Approximately 20%-25% of children with MLD are affected by the juvenile form, which is typically fatal before the age of 20 years.^{5, 6} The adult form of MLD is the least common, with slower progression, characterised by periods of stability and progression continuing until death (typically occurring between 6 and 14 years after diagnosis).^{1, 4} The presenting symptoms of MLD vary by form and include muscle weakness, hypotonia, clumsiness, dysarthria, cognitive regression and neurological issues (weakness and loss of coordination progressing to spasticity and incontinence).¹ Individuals with juvenile or adult forms may present with a decline in school or job performance, behavioural or emotional problems, or psychosis.¹

MLD is usually detected after birth and once symptoms have manifested, unless there is an awareness of family history/mutation status or previous development of MLD in a sibling.^{5, 7}

Potential methods of screening for MLD utilise the measurement of sulfatide levels in urine or dried blood spot (DBS) samples,⁸ the quantification (by immunoassay) of ARSA protein abundance in DBS⁹ and the measurement of ARSA enzymatic activity in DBS samples using tandem mass spectrometry.¹⁰ Studies have assessed sulfatide analysis and ARSA enzymatic activity individually (single tier screening),¹⁰ or in combination as a 2-tier screening strategy.¹¹⁻¹³ The 2-tier screening strategy also has the potential to identify individuals with multiple sulfatase deficiency (MSD), another ultra-rare lysosomal storage disorder where affected patients also display high sulfatide levels and low ARSA enzymatic activity in the blood.^{11, 14} The treatment options for individuals with MSD are limited to management of symptoms and supportive care.¹⁵ For both MLD and MSD, early identification may be useful for reproductive planning, as carrier testing for at-risk family members and prenatal testing for pregnancies at increased risk are possible using molecular genetic techniques if the pathogenic variants in the family are known.¹⁵ Low ARSA enzymatic activity alone is not considered sufficient for the diagnosis of MLD. This is due to the relatively high prevalence of the ARSA pseudodeficiency allele, which leads to reduced enzyme activity (5% to 20% of that of normal controls),¹ but which is not known to manifest as disease or neurological symptoms.¹⁶ Relatively high rates of detection of ARSA pseudodeficiency have also been reported for screening using quantification of ARSA levels by immunoassay.¹⁷ Questions have also been raised about the sensitivity of the immunoassay method as properly folded, but enzymatically deficient proteins could potentially give rise to false negatives.^{11, 18} The thermal instability of ARSA adds a potential logistic challenge in that inadequate sample storage conditions can result in ARSA degradation and hence generate false positives.⁹ Genetic testing is generally recommended to confirm a diagnosis of MLD and

genetic confirmatory testing is considered the reference standard for screening.² Magnetic resonance imaging (MRI) brain scans can also be used to inform a diagnosis of MLD.²

Interventions evaluated for the treatment of MLD have included bone marrow or haematopoietic stem cell transplantation (HSCT), enzyme replacement therapy, cell therapies and gene therapies.¹⁹ However, HSCT has been shown to have limited efficacy and is associated with a significant risk of complications.²⁰ Historically, best supportive care (BSC) and the management of symptoms have been the main focus of treatment, particularly for individuals with late infantile MLD in whom disease management has focussed on palliative care.^{5, 7}

Atidarsagene autotemcel (ARSA-cel/OTL-200, developed by Orchard Therapeutics and branded as Libmeldy®) is recommended by the National Institute for Health and Care Excellence (NICE), Highly Specialised Technology (HST) guidance 18 (HST18), as an option for treating MLD in presymptomatic children with late infantile or early juvenile MLD, and in children with early juvenile MLD who have early clinical signs or symptoms (who can still walk independently and who have no cognitive decline).⁵ Libmeldy® is an autologous haematopoietic stem cell gene therapy (HSC-GT), which involves removing and correcting a patient's stem cells by inserting a functional copy of the ARSA gene, before returning the cells to the patient.⁵ Libmeldy® should be delivered in a highly specialised service by a specialist multidisciplinary team.⁵

The first baby to be treated with Libmeldy® in the UK National Health Service (NHS) was treated at the Royal Manchester Children's Hospital in 2022.²¹ Treatment began with stem cell harvest at 12 months of age and transplant of the treated stem cells took place in August 2022. The patient was discharged home in October 2022 and several months later (February 2023), "has fully recovered from the transplant and is showing no signs of the devastating disease she was born with."²²

Libmeldy® was approved by the European Medicines Agency (EMA) in 2020 and by the U.S. Food and Drug Administration (FDA) in March 2024 for the treatment of presymptomatic late infantile, presymptomatic early juvenile or early symptomatic early juvenile MLD.²³

Current policy context and previous reviews

There is a simple discount Patient Access Scheme (PAS) for Libmeldy® in place in the NHS in England, which is scheduled for review in 2025.⁵

Routine newborn screening for MLD is not currently recommended by the UK National Screening Committee (NSC). Screening was discussed during the appraisal process which informed NICE guidance HST18⁵ where clinical and patient experts highlighted the importance of early diagnosis and newborn blood spot (NBS) screening for inherited disorders such as MLD, and NICE appraisal committee's members acknowledged the difficulties of diagnosis without knowledge of an affected sibling.^{5, 7}

MLD is not included in the Recommended Uniform Screening Panel (RUSP) in the United States (US).²⁴ It was included in the list of conditions nominated to the RUSP (August 2024).²⁵

Screening for MLD has not previously been considered by the UKNSC. It was proposed as a potential NBS population screening programme in the 2021 annual call for topics. The submission reasoned that, without screening, affected individuals are only identified before symptom onset when an older sibling is affected and that this limits the opportunity for treatment

in individuals without affected siblings. In 2023, a preliminary evidence map was commissioned by the UKNSC to evaluate the volume and type of evidence related to newborn screening for MLD. The evidence map⁷ considered the following questions:

- What is the volume and type of evidence on the accuracy of newborn screening strategies for MLD using DBSs?
- What is the volume and type of evidence available on the benefits and/or harms of interventions in presymptomatic/asymptomatic children with MLD identified through screening? i.e. does early initiation of treatment following screening provide better outcomes for MLD compared with initiation of treatment following clinical detection?
- What is the volume and type of evidence on the cost effectiveness of treatment or screening for MLD in asymptomatic or symptomatic patients?

The 2023 UKNSC evidence map included 25 references, the majority of which (19 references) related to the treatment question.⁷ The evidence map included one US study which evaluated a 2-tier screening algorithm (combining quantification of C16:0-sulfatide with measurement of ARSA enzymatic activity) for MLD screening using DBSs from 27,000 newborns. The evidence map also noted that two prospective pilot studies were ongoing in Northern Germany and in New York State, US. For the treatment question, publications relating to 19 cohort and case-control studies were included. The interventions evaluated in these studies included gene therapy (most commonly Libmeldy®, 14 publications), HSCT and umbilical cord blood transplantation. These publications evaluated the efficacy and safety of treatments in presymptomatic patients with MLD and included some comparisons of outcomes with untreated or symptomatic treated patients. However, none of the studies included in the evidence map reported cohorts that were explicitly stated to have been identified through NBS screening or cascade testing, i.e. no studies were identified which could provide information on the relative efficacy of a given treatment in early (screening or cascade testing) versus late (symptomatic clinical detection) diagnosed patients with MLD. Four studies, reported in five conference abstracts, were included for the cost effectiveness question; three studies evaluated the cost effectiveness of treatment with Libmeldy® and one study evaluated the cost effectiveness of NBS screening for MLD.

The evidence map concluded that there was sufficient evidence to justify commissioning an evidence summary and that MLD should be added to the UKNSC's recommendation list, to be kept under regular review. The evidence provided by the evidence map was presented and discussed by the UKNSC in June 2023. The committee agreed with the conclusions of the evidence map and recommended that further work on screening for MLD should be commissioned in the form of a full evidence summary including all the questions examined by the evidence map.⁷

This evidence summary provides a summary the published evidence currently available to assess four key UKNSC criteria:²⁶

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

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

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.

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

In order to maintain relevance to current practice, and for consistency with the 2023 UKNSC evidence map,⁷ we have included relevant literature published since January 2012. It should also be noted that the initial evidence map only gauged the volume and type of evidence available and did not involve an in-depth assessment of the evidence. This evidence summary builds on the level of detail provided by the evidence map by including an in-depth appraisal and synthesis of the included evidence. As part of the development of this evidence summary, an in-depth assessment of the evidence outlined that some references in the evidence maps were conference abstracts and, therefore, these do not meet the inclusion criteria specified for this evidence summary.

Objectives

The overall aim of this project was to assess the volume, type and direction of evidence relevant to newborn screening for MLD in the UK NHS NBS screening programme. The following research questions were defined to address specific project objectives:

1. What is the accuracy of single test and 2-tier NBS screening strategies for MLD, using DBS samples?
2. Does early initiation of treatment following screening lead to improved outcomes for MLD compared to initiation of treatment following clinical presentation?
3. How have modelling studies and CEAs addressed NBS screening for MLD in the era of novel treatments?

In addition to summarising the available evidence to inform the above questions, we:

- conducted a horizon scanning exercise to identify any ongoing studies and recent developments in novel therapies for MLD
- sought information about any implemented NBS screening programmes for MLD that are relevant to the UK context
- sought published clinical guidelines on the management of MLD that are relevant to the UK context

Table 1: Key questions for the evidence summary and relationship to the UKNSC screening criteria

Criterion	Key questions	Studies/Publications Included	
Screening Test			
4	There should be a simple, safe, precise and validated screening test.	What is the accuracy of single test and 2-tier NBS screening strategies for MLD, using DBS samples?	3 ¹¹ 27 28
5	The distribution of test values in the target population should be known and a suitable cut-off level defined and agreed.		
Treatment			
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.	Does early initiation of treatment following screening lead to improved outcomes for MLD compared to initiation of treatment following clinical presentation?	3 ^{29, 30, 40} *

Criterion	Key questions	Studies/Publications Included
Cost effectiveness		
14	The opportunity cost of the screening programme (including testing, diagnosis and treatment, administration, training and quality assurance) should be economically balanced in relation to expenditure on medical care as a whole (value for money). Assessment against this criterion should have regard to evidence from cost benefit and/or CEAs and have regard to the effective use of available resource.	How have modelling studies and CEAs addressed NBS screening for MLD in the era of novel treatments? 1 ³¹

* An additional publication, Fumagalli et al. (2025)⁶⁵ was added following public consultation.

DBS: dried blood spot; CEAs: cost-effectiveness analyses; NBS: newborn blood spot; MLD: metachromatic leukodystrophy; NSC: National Screening Committee; MSD: multiple sulfatase deficiency; UK: United Kingdom

Methods

The current review was conducted by Kleijnen Systematic Reviews Ltd (KSR), in keeping with the UKNSC evidence review process.

All searching was undertaken to the highest standard to meet best practice requirements recommended by the Centre for Reviews and Dissemination (CRD) and the Cochrane Collaboration Handbook.^{32 33}

A sensitive search strategy was developed to retrieve references to studies on MLD. Search strategies were developed specifically for each database and the keywords adapted according to the configuration of each database. Searches combined relevant search terms comprising indexed keywords (e.g. Medical Subject Headings (MeSH) and Emtree) and free text terms appearing in the title and/or abstract of database records. Search terms were identified through

discussion with the review team, by scanning background literature and ‘key articles’ already known to the review team, and by browsing database thesauri.

Only studies conducted in humans were sought. Searches were not limited by language or by publication status (unpublished or published). In order to maintain relevance to current clinical practice and update existing research, searches were date limited from 2012 to present. Conference proceedings and preprints were not included in the search.

The main Embase strategy for each search was independently peer reviewed by a second Information Specialist based on the Canadian Agency for Drugs and Technologies in Health (CADTH) Peer Review checklist.³⁴

Identified references from the bibliographic database searches were downloaded into Endnote bibliographic management software for further assessment and handling. Individual records within the Endnote libraries were tagged with searching information, such as searcher, date searched, database host, database searched, strategy name and iteration, theme or search question. This enabled the Information Specialist to track the origin of each individual database record, and its progress through the screening and review process.

Eligibility for inclusion in the review

The process for selecting studies for inclusion in this evidence summary was as follows:

1. Each title and abstract was reviewed against the inclusion/exclusion criteria by two reviewers, independently. Any disagreements were resolved by discussion and consultation with a third reviewer, as needed.
2. Full-text articles required for the full-text review stage were acquired.
3. Each full-text article was reviewed against the inclusion/exclusion criteria by two reviewers, independently, to determine whether the article was relevant to one or more of the review questions. Any disagreements were resolved by discussion and consultation with a third reviewer, as needed.

Eligibility criteria for each question are presented in Table 2 below. Studies published in languages other than English were excluded. Only studies reported in peer reviewed publications were eligible for inclusion; conference abstracts were excluded.

Table 2: Inclusion and exclusion criteria for the key questions

Key question	Inclusion criteria							Exclusion criteria
	Population	Target Condition	Intervention	Reference standard	Comparator	Outcome	Study type	
1. Screening test	Newborns	MLD	Any screening strategy using DBS samples and single or 2-tier testing.	Confirmatory genetic testing or any specified reference standard.	None or other screening strategy using DBS samples	Sensitivity, specificity, PPV, NPV, of the screening strategy (e.g. by screening test, method of analysis, single or 2-tier testing and threshold) for the target condition MLD. Incidental findings (e.g. MSD).	Studies in randomly assigned or consecutively enrolled populations (diagnostic cohort studies) and diagnostic case-control studies.	None
2. Treatment^a	Newborns, infants or children with MLD.	N/A	Treatment with Atidarsagene (also called Libmeldy®) or any other intervention, where:	N/A	Treatment with Atidarsagene (also called Libmeldy®) or	Survival, symptoms associated with MLD, safety (e.g. incidence of AE associ-	Any comparative study design, in humans, regression analyses	None

			<ol style="list-style-type: none"> 1. MLD has been detected through population screening 2. MLD has been detected in the presymptomatic period (e.g. incidentally or through cascade testing) 		<p>any other intervention, where:</p> <ol style="list-style-type: none"> 1. MLD has been detected without population screening 2. MLD has been detected following symptomatic presentation 	<p>ated with treatment), over-treatment, HRQoL, any other reported outcome.</p>	<p>where treatment outcome is the dependent variable and diagnostic route (e.g. screening/pre-symptomatic detection/symptomatic detection) or time to treatment is an independent variable.</p>	
3. Cost effectiveness	Newborns	MLD	Newborn population screening for MLD.	N/A	No newborn screening for MLD or targeted screening (cascade testing)	Total cost of screening for MLD, incremental cost, incremental life-years gained, gain in any other reported clinical outcome, ICER, number of lives saved, cost per life saved, any	<p>Decision analytic models and economic evaluations.</p> <p>Cost-minimisation, cost effectiveness, cost-utility, cost-benefit and cost-consequence analyses.</p>	None

other reported
outcome
Reviews of
economic
evaluations.

^a If no studies are identified which explicitly compare the efficacy of treatments for MLD in early (screening or cascade testing) versus late (symptomatic presentation) detection, studies comparing the treatment of presymptomatic people with MLD to no treatment (natural history) or treatment of symptomatic MLD, and studies assessing correlation between time to treatment and outcome will be included.

AE: adverse events; DBS: dried blood spot; HRQoL: health-related quality of life; ICER: incremental cost-effectiveness ratio; MLD: metachromatic leukodystrophy; MSD: multiple sulfatase deficiency; N/A: not applicable; NBS: newborn blood spot; NPV: negative predictive value; PPV: positive predictive value

Data extraction

Data were extracted by one reviewer, using piloted data extraction forms. A second reviewer checked the data extraction and any disagreements were resolved by consensus or discussion with a third reviewer.

Appraisal for quality/risk of bias tool

The methodological quality of included studies was assessed by one reviewer and checked by a second reviewer; any disagreements were resolved by discussion or by consultation with a third reviewer. A summary of risk of the methodological quality of included studies is provided in the question level synthesis and full risk of bias assessments, for each study, are provided in Appendix 3.

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

- studies which reported test performance characteristics for one or more screening algorithms for NBS screening for MLD: Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool³⁵ and QUADAS-comparative (QUADAS-C),³⁶ when appropriate
- observational studies which used simple pairwise comparisons with historical natural to assess the effectiveness of treatments in patients with pre-symptomatic/early symptomatic MLD: Risk Of Bias In Non-Randomized Studies – of Interventions (ROBINS-I)³⁷
- cost effectiveness studies: guidelines for authors and peer reviewers of economic submissions to the British Medical Journal (BMJ) (the Drummond checklist)³⁸

Methods of analysis/synthesis

A narrative synthesis of results is presented, structured by UKNSC criterion and key question. No meta-analyses were conducted.

Databases/sources searched

Search strategies were developed to identify studies on MLD, as recommended in the CRD guidance for undertaking reviews in health care³² and the Cochrane Handbook for Diagnostic Test Accuracy Reviews.³⁹

Candidate search terms were identified from target references, browsing database thesauri (e.g. MEDLINE MeSH and Embase Emtree), existing reviews and initial scoping searches. Strategy development involved an iterative approach testing candidate text and indexing terms across a sample of bibliographic databases, aiming to reach a satisfactory balance of sensitivity and specificity. Search strategies were developed specifically for each database and the keywords and thesaurus terms were adapted according to the configuration of each database.

In order to maintain relevance to current clinical practice and to maintain consistency with the UKNSC 2023 Evidence Map, searches carried a date limit of 2012 to present (October 2024).

For details of the full search strategies used please see Appendix 1.

Searches were conducted on the following resources:

- MEDLINE and Epub Ahead of Print, In-Process, In-Data-Review & Other Non-Indexed Citations and Daily (Ovid): 2012-2024/10/21
- Embase (Ovid): 2012-2024/10/21
- Cumulative Index to Nursing and Allied Health Literature (CINAHL) (EBSCO): 2012-2024/10/23
- Cochrane Database of Systematic Reviews (CDSR) (Wiley): 2012-2024/10/Iss10
- Cochrane Central Register of Controlled Trials (CENTRAL) (Wiley): 2012-2024/10/Iss9
- KSR Evidence (KSR) (Internet) (<https://ksrevidence.com/>): up to 2024/10/22
- Orphanet (Internet) (<https://www.orpha.net/en/disease>): up to 2024/10/22
- Orphanet Newborn Screening Bibliographical Knowledgebase (Internet) (<https://nbs.orphanet.app/>): up to 2024/10/22

Horizon scanning searches

Completed and ongoing trials were identified by searches of the following resources:

- National Institutes of Health (NIH) ClinicalTrials.gov (Internet) (<http://www.clinicaltrials.gov/>): up to 2024/10/15
- EU Clinical Trials Register (Internet) (<https://www.clinicaltrialsregister.eu/ctr-search/search>): up to 2024/10/21
- World Health Organization International Clinical Trials Registry Platform (WHO ICTRP) (Internet) (<http://www.who.int/ictrp/en/>): up to 2024/10/15
- ScanMedicine (Internet) (<https://scanmedicine.com/>): up to 2024/10/21

Additional searches

A search of the following resources was conducted to identify background, guideline and policy documents on MLD:

- Trip Database (Internet) (<https://www.tripdatabase.com/>): up to 2024/10/21
- Guidelines International Network (GIN) (Internet) (<https://g-i-n.net/international-guidelines-library/>): up to 2024/10/21
- NICE (<https://www.nice.org.uk/>): up to 2024/10/21
- International HTA Database (Internet) (<https://database.inahta.org/>): up to 2024/10/22
- National Institute for Health and Care Research (NIHR) Health Technology Assessment (HTA) (Internet) (<https://www.nihr.ac.uk/>): up to 2024/10/21

- Europe, Middle East & Africa (ECRI) Guidelines Trust (Internet) (<https://home.ecri.org/>): up to 2024/10/21

Update searches

In order to identify any relevant primary studies published since the original strategies were run in October 2024, the main Embase and MEDLINE searches were rerun in their entirety in January 2025. Results were deduplicated against the original search results and for completeness the medRxiv preprints database was also searched for any relevant forthcoming papers, limiting the date to those papers “posted between " 01 Jan, 2021 and 03 Feb, 2025”:

- MEDLINE and Epub Ahead of Print, In-Process, In-Data-Review & Other Non-Indexed Citations and Daily (Ovid): 2012-2025/01/29
- Embase (Ovid): 2012-2025/01/29
- medRxiv (Internet): 2021-2025/02/03

The main Embase strategy for each search was independently peer reviewed by a second Information Specialist based on the CADTH Peer Review checklist.³⁴

Overview of included studies

The literature searches conducted for this evidence summary identified 2,547 unique publications, after deduplication. Following initial screening of titles and abstracts, 38 publications were considered to be potentially relevant and ordered for full paper screening; of these, seven are included in the Question level synthesis.^{11, 27-31, 40} The update searches identified 127 unique publications, after deduplication, all of which were excluded at the first stage of inclusion assessment (title and abstract review).

Three publications provided data to inform research question 1 on test accuracy,^{11, 27, 28} three publications^{29, 30, 40} relating to two studies provided data to inform research question 2 on treatment, and one publication provided data to inform research question 3 on cost-effectiveness.³¹

Three of the publications included in this evidence summary^{11, 29, 40} had previously been identified by the 2023 UKNSC evidence map.⁷

Appendix 2 provides a Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) flow chart for this evidence summary and details of studies included and excluded after full-text screening. In addition, Appendix 2 includes details of the studies included in the 2023 UKNSC evidence map,⁷ whether these studies are included in the current evidence summary and reasons for exclusion (where applicable).

In addition to the systematic literature review, this evidence summary included a horizon scanning exercise; no ongoing studies of novel therapies for MLD were identified.

We also sought information about any existing implemented NBS screening programmes for MLD that are relevant to the UK context, irrespective of whether such programmes were associated with published evaluations. The 2023 UKNSC evidence map,⁷ noted that two prospective pilot studies were ongoing in Northern Germany and in New York State, US, and this evidence summary includes one publication relating to the German pilot study.²⁷ One publication,¹³ which did not meet the inclusion criteria for this evidence summary (excluded at the full-text screening stage, see Table 38, Appendix 2), listed the following MLD newborn screening pilot studies initiated in the 2 years prior to 2024:

- An assay validation study at the Hospices Civils de Lyon, France.
- A pre-pilot study of 3,687 newborns at the Royal Manchester Children's Hospital (Manchester Biochemical Genetics Laboratory, UK) – related publication included in this evidence summary.²⁸
- A prospective pilot study at Rouen University Hospital in Rouen, France (50,000 newborns to date).
- Three prospective pilot studies by Archimedlife in two regions in Germany and nationwide in Austria (now at >150,000 newborns) – related publication included in this evidence summary.²⁷
- A prospective pilot study at the Meyer's Children's Hospital in Florence, Italy (started in March 2023).

- A prospective pilot study at King Fahad Medical City in Saudi Arabia (~3,000 newborns).

and further noted that, in the US, a multiplex study (ScreenPlus) including MLD, was initiated in 2021, and in 2023, the Illinois legislature approved a Bill to add MLD to the state newborn screening panel, and implementation is anticipated to start in 2024/25. Future publication of data from these studies has the potential to provide evidence to inform criteria 4 and 5.

With respect to implemented NBS screening programmes for MLD, we identified a news article from Oslo University Hospital reporting that: *'In January 2025, Norway became the first country in the world to start national screening for metachromatic leukodystrophy (MLD).'*⁴¹ We were not able to identify any further published details about the new Norwegian screening programme. A 2024 landscape assessment of newborn screening in Europe⁴² includes only one entry for MLD (a pilot programme reported for Italy); although no source for this information was cited it should be noted we identified a separate reference to a prospective pilot study at the Meyer's Children's Hospital in Florence, Italy (see above)¹³ and that the Meyer's Children's Hospital also participated in the feasibility study, Hong et al. (2021)¹¹ included in this evidence summary (providing DBS samples from patients with confirmed MLD).

Finally, guidelines searches were undertaken to identify published clinical guidelines on the management of MLD that are relevant to the UK context. Guidelines searches and a review of our main searches identified two relevant publications, NICE guidance HST18: 'Atidarsagene autotemcel for treating metachromatic leukodystrophy',¹⁵ and a journal article reporting development of a clinical guideline: 'Newborn screening in metachromatic leukodystrophy – European consensus-based recommendations on clinical management.'⁴³ Recommendations from these two publications are summarised in Appendix 5.

Question level synthesis

Criteria 4 and 5 - Accuracy of the screening test

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

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

Question 1 — What is the accuracy of single test and 2-tier NBS screening strategies for MLD, using DBS samples?

The 2023 UKNSC evidence map⁷ identified one study, Hong et al. (2021)¹¹ which was relevant to the question:

What is the volume and type of evidence on the accuracy of newborn screening strategies for MLD using DBSs?

The 2023 UKNSC evidence map noted that Hong et al. (2021)¹¹ evaluated a 2-tier screening strategy for MLD, followed by genetic confirmatory testing for clinically relevant *ARSA* variants, which was designed to have 100% sensitivity and was subsequently found to have almost 100% specificity.⁷ However, it was further noted that: *'at present, evidence on the performance of the 2-tier newborn screening strategy in DBS samples for MLD is limited'* and that two ongoing pilot studies in Northern Germany⁴⁴ and New York¹² may contribute further evidence on this screening strategy.⁷ No studies were identified which met the eligibility criteria for the 2023 UKNSC evidence map and evaluated single screening tests (measurement of sulfatides or *ARSA* enzymatic activity) followed by genetic confirmatory testing; the 2023 UKNSC evidence map stated that further work to evaluate all available screening strategies would be justified.⁷

What is added by this evidence summary

This evidence summary provides a summary of the published studies available to inform question 1, which includes Hong et al. (2021)¹¹ (included in the 2023 UKNSC evidence map⁷), a recent published report from a German pilot study,²⁷ (identified as an ongoing study by the 2023 UKNSC evidence map⁷), and a report of a newly identified UK pre-pilot study.²⁸ We did not identify any publications relating to the ongoing US pilot study¹² noted by the 2023 UKNSC evidence map.⁷

No studies were identified which reported experience from implemented screening programmes.

Description of new evidence in relation to previous evidence reviews

The searching and title and abstract screening stages of the evidence summary were conducted as a single process, with consideration of all three research questions. Appendix 2 provides an overall PRISMA flow chart for this evidence summary and details of studies included and excluded after full-text screening.

Following full-text screening, there were three publications that met the inclusion criteria specified for research question 1.^{11, 27, 28} One publication, a report of a prospective feasibility

study,¹¹ was included in the 2023 UKNSC evidence map.⁷ The two new publications identified by this evidence summary reported results from a prospective pilot study, Laugwitz et al. (2024),²⁷ conducted in Germany and a prospective 'pre-pilot' study, Wu et al. (2024),²⁸ conducted in the UK; both of these studies were partially funded by Orchard Therapeutics, the company which manufactures Libmeldy® (recommended for the treatment of pre-symptomatic and early symptomatic MLD).⁵ The funders had no role in study design or experimental implementation of these studies. The German pilot study was published in the *New England Journal of Medicine* as a letter to the editor, with a detailed study report provided as supplementary material;²⁷ although not a peer-reviewed publication the article has been included in this evidence summary in the interests of providing the fullest information possible.

All three of the publications included in this section evaluated 2-tier screening strategies involving the measurement of sulfatides and ARSA enzyme activity;^{11, 27, 28} Laugwitz et al. (2024)²⁷ described a 3-tier algorithm where 'confirmatory' genetic testing of the DBS sample comprised the 3rd tier. A summary of the screening algorithms evaluated by the included studies is provided in Table 3 and illustrative flow charts are provided in Appendix 4. We did not identify any studies that evaluated single screening tests.

Feasibility study

The feasibility study, Hong et al. (2021),¹¹ used DBS samples from de-identified newborns to evaluate a 2-tier screening algorithm, which had been designed to minimise the false positive rate (FPR). The 1st tier of the screening algorithm comprised the quantification of C16:0-sulfatide in DBS, using ultraperformance liquid chromatography – tandem mass spectrometry (UPLC-MS/MS) and the 2nd tier comprised an ARSA enzymatic activity assay, using a tandem mass spectrometry method previously published by the authors.¹⁰

The DBS samples from de-identified newborns used in this study were provided by the Washington State Department of Health, US, after being stored at room temperature for 30 to 60 days. A total of 15 archived DBS samples from newborns with confirmed MLD were acquired through the MLD foundation, University of Pittsburgh, US and the Meyer Children's Hospital, Florence, Italy, and were used to establish reference ranges; these samples had been stored at -20°C.¹¹

In the first phase of the Hong et al. (2021)¹¹ study, the 1st tier screening cut-off was established based on C16:0-sulfatide concentrations in DBS from 15 newborns with confirmed MLD (median 0.32 µmol/L, range 0.18 to 0.47 µmol/L) and 2,000 random newborns (median 0.094 µmol/L, range 0.020 to 0.23 µmol/L). With consideration to the overlap between these distributions, the screening cut-off required to achieve 100% sensitivity was judged to be ≥0.17 µmol/L.¹¹ The final method used for 1st tier screening was quantification of C16:0-sulfatide in DBS using d₅-C16:0-sulfatide as the internal standard and normalised to the external calibrator (14.4 nmol/L C16:0-sulfatide in methanol). Consideration of the normalised sulfatide levels in the DBS from six (insufficient residual sample for the remaining nine) of the newborns with confirmed MLD (median 1.24, range 0.68 to 1.48) and the DBS from 2,000 random newborns (median 0.34, range 0.11 to 0.86) resulted in the definition of the cut-off needed to achieve 100% sensitivity as ≥0.64, after normalisation.¹¹

Measurement of ARSA enzymatic activity was implemented for 2nd tier screening.¹¹ The enzymatic activity assay used a tandem mass spectrometry method, previously published by the authors,¹⁰ and a second 3 mm punch from the same DBS used for the sulfatide analysis. To account for the thermal instability of ARSA, the reference range had previously been

established using DBS samples from newborns with confirmed MLD and ‘matching newborns’ (newborns with normal sulfatide levels with similar sample storage conditions); the 2nd tier screening cut-off was set at <20% of the mean activity for ‘matching newborns’.¹⁰ If ARSA enzymatic activity was low, the activities of three additional sulfatases (iduronate-2-sulfatase [I2S], N-Acetylgalactosamine-6-sulfatase [GALNS] and Arylsulfatase B [ARSB]) were measured to distinguish MLD screen positives from MSD screen positives.¹¹

In the second phase of the Hong et al. (2021) study, 27,335 DBS from random newborns were screened using the strategy described above, of which 195 (0.71%) had a normalised C16:0-sulfatide level above the cut-off.¹¹ Of the 195 DBS samples with C16:0-sulfatide levels above 0.64, after normalisation, 122 entered 2nd tier screening; the remaining 73 were not tested for ARSA enzymatic activity because the DBS samples were considered to be too old (stored at room temperature for >3 months) and were excluded from analyses. All but two of the tested samples had ARSA enzymatic activity >20% of the mean for ‘matching newborns’, i.e. they were 2nd tier screen negative and hence screen negative for MLD.¹¹

One of the MLD screen positive samples had a normalised C16:0-sulfatide level of 0.86 and 0% ARSA enzymatic activity, and the second MLD screen positive sample had a normalised C16:0-sulfatide level of 0.72 and 8% ARSA enzymatic activity. I2S, GALNS and ARSB activities were only measured in the second sample and all were >20% of the mean, i.e. negative for MSD. ARSA exome sequencing was carried out on a third 3 mm punch from each of these DBS samples; sequencing results for the first sample was interpreted as an MLD-affected patient and the second sample was interpreted as a heterozygote (carrier); the estimated FPR for this screening algorithm was 0.0037%.¹¹

No routine further analyses, records review or surveillance was reported, either in relation to the whole MLD screen negative population (n=27,333) or the 195 1st tier screen positive DBS samples; although the authors noted that future surveillance would be possible, since the approximate birthdates of newborns tested were known.¹¹ ARSA exome sequencing was undertaken, and no pathogenic variants detected for 3/122 1st tier screen positive DBS samples, which had elevated sulfatide levels and normal ARSA enzymatic activity, however, it was not clear why these samples were selected for sequencing.¹¹ Given the absence of any reported method for identifying MLD cases that may have been missed by screening, both the true sensitivity achieved by the proposed screening algorithm and the prevalence of MLD in the tested population remain subject to uncertainty. If it is assumed that no cases of MLD were missed and participants whose DBS did not receive 2nd tier testing are excluded from the analysis (i.e. the numbers of true positives [TP], false positives [FP], false negatives [FN] and true negatives [TN] were 1, 1, 0 and 27,260, respectively), then the estimated positive predictive value (PPV) for the screening algorithm would be 50% (95% confidence interval [CI]: 12.35 to 87.65%) and the negative predictive value (NPV) would be 100%.

The Hong et al. (2021) study also explored the possibility of using ARSA enzymatic activity in DBS as the 1st tier screening test for MLD, with the C16:0-sulfatide assay implemented as the 2nd tier test. This approach was evaluated using de-identified DBS samples from 2,287 newborns. Three of these samples had ARSA enzymatic activity below cut-off, all of which were considered MLD screen negative based on the results of C16:0-sulfatide analysis; no further confirmatory testing was reported.¹¹

Measures of the screening algorithm performance are summarised in Table 4 and details of clinically relevant variants identified are provided in Table 5.

Prospective pilot and pre-pilot studies

Two further studies, a large (n=109,259) pilot study conducted in Germany²⁷ and a small (n=3,687) 'pre-pilot' study conducted in the UK,²⁸ evaluated screening algorithms for MLD, which were based on the methods described in Hong et al. (2021).¹¹

The large pilot study by Laugwitz et al. (2024) included neonates in hospitals referring to the newborn screening laboratory in Hannover, Germany, between October 2021 and July 2023, in whom DBS had been collected for NBS within the first 36 to 72 hours of life (following the national guidelines for NBS in Germany) and for whom residual DBS sample was available after completion of the regular national NBS programme and legal guardian consent had been obtained.²⁷ The study evaluated a 3-tier screening algorithm where the 1st tier used the UPLC-MS/MS assay described in Hong et al. (2021),¹¹ but utilising measurement of two sulfatides and a cut-off of $\geq 0.17 \mu\text{mol/L}$ cut-off for C16:0-sulfatide **or** C16:1-OH-sulfatide $\geq 0.05 \mu\text{mol/L}$, and the 2nd tier used the tandem mass spectrometry assay of ARSA enzymatic activity, described in Hong et al. (2021),¹¹ but with an absolute rather than a relative cut-off ($\leq 0.015 \mu\text{mol/L/h}$). The 3rd tier screening test was genetic sequencing, using the DBS sample, to identify clinically relevant variants in *ARSA*, *SUMF1* or *PSAP*, a step which might more usually be regarded as part of confirmatory testing and which comprised the confirmatory testing method in Hong et al. (2021).¹¹ Laugwitz et al. (2024) described confirmatory diagnosis at a qualified treatment centre, for MLD screen positive newborns, as comprising measurement of ARSA in blood and sulfatides in urine, and confirmatory genetic sequencing of samples from newborn and parents.²⁷

Using this algorithm, 381/109,259 (0.35%) of the newborns screened were 1st tier positive (i.e. had at least one sulfatide, C16:0 **and/or** C16:1-OH above the cut-off). Due to (unspecified) technical issues, ARSA enzymatic activity testing was only carried out in 230/381 1st tier positive DBS; 20 samples showed enzymatic activity below the cut-off (i.e. were 2nd tier positive), giving an overall 2nd tier positive rate of 0.018%.²⁷

Third tier genetic testing of DBS samples was conducted in all 381 samples with elevated sulfatides (i.e. in all 1st tier screen positive samples, irrespective of whether 2nd tier screening was positive, negative or not undertaken). Three samples, each with two presumed compound heterozygous pathogenic *ARSA* variants, were classified as MLD screen positive; all three of these samples had C16:0-sulfatide **and** C16:1-OH-sulfatide levels above the cut-off and no measurable ARSA enzymatic activity, and confirmatory biochemical and genetic testing confirmed the diagnosis of MLD in all three cases. Six further samples had only one clinically relevant *ARSA* variant and were identified as screen negative heterozygous carriers. Finally, 4 samples with one clinically relevant Prosaposin (*PSAP*) variant and 3 samples with one clinically relevant sulfatase modifying factor 1 (*SUMF1*) variant (i.e. 4 screen negative *PSAP* carriers and 3 screen negative *SUMF1* carriers) were identified; no samples were screen positive for MSD. If 3rd tier testing (genetic sequencing of DBS samples) is treated as part of the confirmatory testing process and the heterozygous carriers (n=13) as well as the 4 1st and 2nd tier positive patients for whom no genetic testing results were reported are treated as negative for MLD on confirmatory testing, the estimated overall FPR for 1st and 2nd tier testing (i.e. a 2-tier screening algorithm) is 0.016%.²⁷

With respect to the identification of potential screening FN cases, genetic sequencing of the 210 DBS samples that were 1st tier positive and 2nd tier negative (i.e. had sulfatide levels above the cut-off, but ARSA enzymatic activity within the normal range) and the 151 1st tier positive DBS samples for which no 2nd tier testing was undertaken, did not identify any further MLD cases.²⁷

No routine further analyses, records review or surveillance was reported in relation to the screen negative population.²⁷ Given the absence of any reported comprehensive method for identifying MLD cases that may have been missed by screening, both the true sensitivity achieved by the proposed screening algorithm and the prevalence of MLD in the tested population remain subject to uncertainty. If it is assumed that no cases of MLD were missed and participants whose DBS did not receive 2nd tier testing are excluded from the analysis (i.e. the numbers of TP, FP, FN and TN were 3, 17, 0 and 109,088, respectively), then the estimated PPV for a 2-tier screening algorithm, derived from Laugwitz et al. (2024), would be 15% (95% CI: 9.89 to 22.11%) and the NPV would be 100%.

Based on the *ARSA* genotype, early juvenile disease onset was predicted for two of the three confirmed MLD cases and late juvenile or adult onset was predicted for the remaining case. The two infants with predicted early juvenile disease onset were treated with myeloablative chemotherapy (intravenous [IV] busulfan) before administration of *ARSA*-cel at age 12 months; at last follow-up (age 18 months) both infants had reached all developmental milestones and had 'unremarkable' MRI.²⁷ The infant with predicted late juvenile or adult disease onset was scheduled for yearly clinical monitoring with HSCT planned between the age of 2 and 5 years; at the time of last follow-up (age 15 months) all developmental milestones had been achieved.²⁷

The small UK 'pre-pilot' study, Wu et al. (2024) was conducted at the Centre for Genetic Medicine, Manchester University NHS Foundation Trust, to assess the feasibility of using a 2-tier algorithm for newborn screening for MLD in the UK.²⁸ The study used de-identified residual DBS from the established UK newborn DBS screening programme; DBS samples were excluded if they were collected from babies ≤ 4 days or > 12 months of age, rejected due to blood transfusion or poor sample quality, or if parents declined any research being performed on the baby's residual sample.²⁸ The study evaluated a 2-tier screening algorithm where the 1st tier used the UPLC-MS/MS assay of C16:0-sulfatide levels and the 2nd tier used the tandem mass spectrometry assay of *ARSA* enzymatic activity, described in Hong et al. (2021).¹¹

A validation phase was undertaken before the 'pre-pilot' study to evaluate the published¹¹ sulfatide and *ARSA* enzymatic activity cut-offs. C16:0-sulfatide was evaluated using DBS from known MLD patients (n=13), unaffected siblings age < 12 months (n=3) and patients with *ARSA* pseudodeficiency age < 12 months (n=4).²⁸ All DBS from MLD patients had sulfatide levels above the published cut-off (≥ 0.17 $\mu\text{M/L}$) and all unaffected siblings and patients with *ARSA* pseudodeficiency had sulfatide levels below the cut-off. 657 Newborn DBS were used to establish a neonatal reference range for C16:0-sulfatide, 0.045 to 0.215 $\mu\text{mol/L}$ (mean = 0.09 $\mu\text{mol/L}$) leading to 3 (0.4%) positive results; based on these initial assessments, the C16:0-sulfatide ≥ 0.17 $\mu\text{M/L}$ cut-off was deemed appropriate for 1st tier screening. *ARSA* enzymatic activity was evaluated, using DBS stored at -20°C for up to 6 months or residual NBS within 8 days after heel prick, from known MLD patients (n=12), unaffected siblings age < 12 months (n=4), patients with *ARSA* pseudodeficiency (n=8) and known MLD-negative DBS (n=46). *ARSA* enzymatic activity of $< 20\%$ of negative controls¹¹ correctly identified all (n=12) DBS from known MLD patients and all known MLD negative DBS (normal n=46, pseudodeficiency n=8, unaffected siblings n=4) had quantifiable *ARSA* enzymatic activity $> 20\%$ of the negative controls. A UK *ARSA* enzymatic activity reference range was established using 120 newborn DBS, 0.042 to 0.689 $\mu\text{mol/L/h}$ (mean 0.333 $\mu\text{mol/L/h}$), with two newborns having *ARSA* enzymatic activity $< 20\%$ of the mean.

The 'pre-pilot' study²⁸ included 3,687 DBS samples and evaluated a 2-tier screening algorithm using the published cut-off values for C16:0-sulfatide and *ARSA* enzymatic activity of ≥ 0.17 $\mu\text{M/L}$ and $< 20\%$ of normal mean, respectively.¹¹ Using this algorithm, 11/3,687 (0.3%) of

samples were 1st tier screen positive, all of which had ARSA enzymatic activity $\geq 20\%$ of the normal mean (i.e. the 'pre-pilot' study did not identify any MLD screen positive samples). ARSA gene sequencing was conducted on all 11 1st tier screen positive DBS samples and no pathogenic variants were identified.²⁸

Of the two samples from the validation phase, with ARSA enzymatic activity $< 20\%$ of the mean, both were subsequently found to have sulfatide levels below the $\geq 0.17 \mu\text{M/L}$ cut-off ($0.15 \mu\text{mol/L}$ and $0.086 \mu\text{mol/L}$) and both were submitted for ARSA gene sequencing. Genetic sequencing of the DBS which had a C16:0-S level of $0.15 \mu\text{mol/L}$ and ARSA enzymatic activity of 4% of mean revealed that the sample was homozygous for c.465+1G>A, a common pathogenic variant associated with late infantile MLD (confirmed by repeat testing of the sample). The child was 10 months old, below the predicted age of onset for late infantile disease (18-24 months) and therefore likely to be pre-symptomatic and eligible for ARSA-cel gene therapy. Following an urgent case review, permission was given for the sample to be de-anonymised and the child was referred for specialist assessment. The MLD diagnosis was confirmed by deficient leukocyte ARSA activity (5.8 nmol/mg/h , normal range $45\text{-}250 \text{ nmol/mg/h}$) and trio testing showing inheritance of the pathogenic variant from both parents. The child commenced ARSA-cel gene therapy at 11 months old and remained under review and symptom free at 19 months old.²⁸

If the C16:0-S cut-off were lowered to $0.15 \mu\text{mol/L}$ the 1st tier positive rate would increase to 0.76%; ARSA enzymatic activity was not assessed in these additional 17 samples due to sample storage conditions. The study used the results of additional testing, retrospective C16:1-OH testing with all samples being below the MLD-positive range, to infer that lowering of the 1st tier threshold would have been unlikely to result in additional screening false positives.²⁸

No further analyses, records review or surveillance was reported in relation to the screen negative population.²⁸

Measures of the screening algorithm performance are summarised in Table 4 and details of clinically relevant variants identified are provided in Table 5.

Methodological quality of studies

It is important to note that although information provided in some publications included in this section has been used to estimate measures of screening performance (PPV and NPV), these publications do not describe diagnostic test accuracy studies intended to evaluate the diagnostic performance of screening tests or algorithms. No study reported the universal application of a diagnostic reference standard; confirmatory genetic testing was only carried out where there was an abnormal (MLD screen positive) result or, in some cases, where only the 1st tier test in a 2-tier screening algorithm was positive. Although the large sample sizes needed to evaluate newborn screening for MLD may mean that genetic sequencing of all samples is not considered practicable, it would be theoretically possible to apply a standardised approach to surveillance for missed cases (FN) as has been demonstrated in other NBS studies, for example in Spinal Muscular Atrophy (SMA) screening.^{45, 46}

QUADAS-2 and, where applicable, QUADAS-C has been applied to both studies from which measures of screening performance could be estimated. The use of QUADAS-2 was considered to be appropriate because the question under consideration (*Question 1 - What is the accuracy of single test and 2-tier NBS screening strategies for MLD, using DBS samples?*) is one of test accuracy; it is therefore important to consider the methodological limitations of the

included studies in respect of their ability to address this question, irrespective of study design/primary aim.

Table 6 provides a summary of the QUADAS-2/QUADAS-C assessments for the two publications from which PPV and NPV estimates were calculated and the corresponding full QUADAS-2 assessments are provided in Appendix 3.^{11, 27} Both of these publications reported derivation of or adjustment to screening cut-offs during the study and both excluded samples for which 2nd tier screening was not undertaken from any analyses. Importantly, given that application of the diagnostic reference standard or long-term follow-up of all screen-negative babies is unlikely to be considered practicable, neither publication reported a standardised approach to identifying and recording any cases missed by screening; Hong et al. (2021) noted that future surveillance would be possible, since the approximate birthdates of newborns tested were known.¹¹ Hence, although Hong et al. (2021) reported that screening cut-offs had been established to achieve 100% sensitivity, the true sensitivity that could be achieved if the proposed screening algorithm were implemented remains uncertain. This is highlighted by the fact that the incidental late infantile MLD case identified by Wu et al. (2024)²⁸ would not have been detected using the cut-offs for 1st tier sulfatide testing established by Hong et al. (2021).¹¹

QUADAS-C has been applied to Hong et al. (2021), for completeness, However, this is of limited informative value, since there are insufficient data reported to determine the accuracy of either screening algorithm or to compare performance between them. It should be noted that the evaluation of the second screening algorithm (ARSA enzymatic activity in DBS as the 1st tier screening test for MLD and C16:0-sulfatide levels as the 2nd tier test) reported in this study was exploratory only and did not include any confirmatory (reference standard) testing.¹¹

Table 3: Summary of the screening algorithms evaluated in included studies

Study	Country	Study type	Description of screening algorithm	Description of follow-up/confirmatory testing
Hong 2021 ¹¹	US and Italy	Feasibility study	<p><u>Algorithm a)</u></p> <p>Sample: 3 mm Punch from a DBS (extracted in methanol containing the internal standard d5-C16:0-sulfatide)</p> <p>1st Tier: UPLC-MS/MS analysis of sulfatide, C16:0-sulfatide <0.17 µmol/L (<0.64 after normalisation to an external standard) = screen negative, C16:0-sulfatide ≥0.17 µmol/L (≥0.64 after normalisation to an external standard) = 1st tier positive</p> <p>2nd Tier: For 1st tier positive samples, ARSA enzyme activity measured in a 2nd (using a previously published protocol¹⁰), ARSA enzyme activity ≥20% matching newborns^a = screen negative, for ARSA activity <20% matching newborns,^a activities of three additional sulfatases (I2S, GALNS and ARSB) measured to distinguish MLD from MSD, all three additional sulfatides ≥20% matching newborn = screen positive for MLD, activity of one or more additional sulfatides <20% matching newborn = screen positive for MSD</p> <p><u>Algorithm b)</u></p> <p>Sample: 3 mm Punch from a DBS (extracted in methanol containing the internal standard d5-C16:0-sulfatide)</p> <p>1st Tier: ARSA enzyme activity (using a previously published protocol¹⁰), ARSA enzyme activity ≥20% daily mean activity = screen negative, <20% daily mean activity = 1st tier positive</p> <p>2nd Tier: For 1st tier positive samples, UPLC-MS/MS analysis of sulfatide, C16:0-sulfatide <0.17 µmol/L (<0.64 after normalisation to an external standard) = screen nega-</p>	<p>ARSA and <i>SUMF1</i> whole exome sequencing using DBS sample^b</p> <p>None reported</p>

			<p>tive, for C16:0-sulfatide $\geq 0.17 \mu\text{mol/L}$ (≥ 0.64 after normalisation to an external standard), activities of 3 additional sulfatases (I2S, GALNS and ARSB) measured to distinguish MLD from MSD, all three additional sulfatides $\geq 20\%$ matching newborn = screen positive for MLD, activity of one or more additional sulfatides $< 20\%$ matching newborn = screen positive for MSD</p>	
Laugwitz 2024 ²⁷	Germany	Pilot study	<p>Sample: 3.2 mm punch from a DBS sample</p> <p>1st Tier: UPLC-MS/MS analysis of sulfatides (using the methods described in Hong 2021¹¹), C16:0-sulfatide $< 0.17 \mu\text{mol/L}$ (1.83 MoM) and C16:1-OH sulfatide $< 0.050 \mu\text{mol/L}$ (3.13 MoM) = screen negative, C16:0-sulfatide $\geq 0.17 \mu\text{mol/L}$ (1.83 MoM) or C16:1-OH sulfatide $\geq 0.050 \mu\text{mol/L}$ (3.13 MoM) = 1st tier positive</p> <p>2nd Tier: For 1st tier positive samples, ARSA enzyme activity (using a previously published protocol¹⁰), ARSA activity $> 0.015 \mu\text{mol/L/h}$ = screen negative, ARSA activity $\leq 0.015 \mu\text{mol/L/h}$ = 2nd tier positive</p> <p>3rd Tier:^c For 2nd tier positive samples, genetic sequencing (<i>ARSA</i>, <i>SUMF1</i>, <i>PSAP</i>) using DBS sample, no compatible <i>ARSA</i> genotype = screen negative, compatible <i>ARSA</i> genotype = screen positive (Clinically relevant variants in <i>ARSA</i> were reported if detected as one homozygous or a combination of two heterozygous variants)</p>	Confirmatory diagnosis at a qualified treatment centre (measurement of ARSA enzyme activity in blood and sulfatides in urine, confirmatory genetic sequencing of samples from newborn and parents)
Wu 2024 ²⁸	UK	Pre-pilot study	<p>Sample: Residual, de-identified newborn screening DBS (no further details reported)</p> <p>1st Tier: UPLC-MS/MS analysis of sulfatides (using the methods described in Hong 2021¹¹), C16:0-sulfatide $< 0.17 \mu\text{mol/L}$ = screen negative, C16:0-sulfatide $\geq 0.17 \mu\text{mol/L}$ = 1st tier positive</p> <p>2nd Tier: For 1st tier positive samples, ARSA enzyme activity (using a previously published protocol¹⁰), ARSA enzyme activity $\geq 20\%$ normal mean activity^d = screen negative; for ARSA activity $< 20\%$ normal mean activity^d,</p>	PCR amplification of <i>ARSA</i> gene from DBS using an automated EZ1 DNA Tissue Kit (QIAGEN) followed by Sanger sequencing ^e

measure reference enzymes sulphamidase and β -galactosidase, both reference enzymes low = poor quality, request repeat sample, at least one reference enzyme normal = screen positive (suspected MLD), refer to specialist LSD clinician

^a Samples from six to eight newborns with normal sulfatide levels and similar storage conditions.

^b Undertaken in two MLD screen positive samples and in 3/120 MLD screen negative samples that were 1st tier positive (elevated sulfatides with normal ARSA enzyme activity).

^c 'Confirmatory' genetic testing classified as 3rd tier of the screening algorithm.

^d Calculated from five to seven NBS bloodspots with normal sulfatide levels and similar storage conditions to normalise for ARSA enzyme activity loss in storage.

^e Undertaken in 11 screen negative samples that were 1st tier positive (elevated sulfatides with normal ARSA enzyme activity) and in two samples from the initial validation of the ARSA activity reference ranges which had ARSA enzyme activity <20% of the mean.

ARSA: Arylsulfatase A; ARSB: Arylsulfatase B; DBS: dried blood spot; GALNS: N-Acetylgalactosamine-6-sulfatase; I2S: iduronate-2-sulfatase; LSD: lysosomal storage disease; MLD: metachromatic leukodystrophy; MoM: multiples of mean; MSD: multiple sulfatase deficiency; NBS: newborn blood spot; *PSAP*: Prosaposin; *SUMF1*: sulfatase modifying factor 1; UPLC-MS/MS: ultraperformance liquid chromatography – tandem mass spectrometry; UK: United Kingdom; US: United States

Table 4: Summary of screening algorithm performance

Study	Study period	Newborns screened, n	1 st Tier positive, n (%)	2 nd Tier positive, n (%)	1 st Tier FPR, %	2 nd Tier FPR, %	Screen-detected MLD, n	Incidence of screen-detected MLD
Hong 2021 ¹¹	N/A (archived samples)	Algorithm a) 27,335	195 (0.71)	2 (0.0073)	0.71 ^{a,b}	0.0037 ^c	1	1 in 27,335 (3.66/100,000) ^c
		Algorithm b) 2287	3 (0.13)	0 (0)	0.13 ^a	NR	0	NE
Laugwitz 2024 ²⁷	Oct 2021 to Jul 2023	109,259	381 (0.35)	20 (0.018)	0.35	0.016 ^{c,d}	3	1 in 36,420 ^c (2.75/100,000) ^c
Wu 2024 ²⁸	N/A (residual de-identified samples)	3,687	11 (0.3)	0 (0)	0.30	0	0	NE

^a Reported.

^b 23 samples that did not receive 2nd tier testing were excluded.

^c Calculated.

^d 151 samples that did not receive 2nd tier testing were excluded from the calculation.

FPR: false positive rate; MLD: metachromatic leukodystrophy; N/A: not applicable; NE: not estimable; NR: not reported

Table 5: Details of clinically relevant variants identified in screen-positive study participants

Study	MLD cases	ARSA variant carriers	Other (non-ARSA) clinically relevant variants
Hong 2021 ¹¹	Total, n=1 1 (heterozygous for the pathogenic variants c.1283C>T and c.1292A>C and heterozygous for the common pseudodeficiency variant c.1178C>G)	Total, n=1 1 (heterozygous for c.1174C>T, heterozygous for the pseudodeficiency variant c.1178C>G and homozygous for the pseudodeficiency variant c.1055A>G)	NR
Laugwitz 2024 ²⁷	Total, n=3 1 (heterozygous for the pathogenic variants c.412C>T and c.1283C>T) 1 (heterozygous for the pathogenic variants c.465+1G>A and c.1283C>T) 1 (heterozygous for the pathogenic variants c.465+1G>A and c.542T>G)	Total, n=6^a c.931G>A, n=1 c.542T>G, n=2 c.991G>A, n=1 c.1283C>T, n=1 c.170G>T, n=1	Total, n=7 <i>PSAP</i> carrier, n=4 <i>SUMF1</i> carriers, n=3
Wu 2024 ²⁸	0 ^b	0	NR

^a Of which three were 1st and 2nd tier positive, one was 1st tier positive and 2nd tier negative, and one did not receive 2nd tier testing.

^b Additional testing of two samples from the assay validation phase, which had ARSA enzyme activity <20% of the mean, resulted in the identification of one MLD case (homozygous for the pathogenic variant c.465+1G>A)

ARSA: Arylsulfatase A; MLD: metachromatic leukodystrophy; NR: not reported; *PSAP*: Prosaposin; *SUMF1*: sulfatase modifying factor 1

Table 6: Summary of QUADAS-2 and QUADAS-C evaluations

Study	Risk of bias			Flow and timing	Applicability concerns		
	patient selection	Index test	Reference standard		patient selection	Index test	Reference standard
Hong 2021 (algorithm a) ¹¹	✓	?	?	✗	✓	✓	?
Hong 2021 (algorithm b) ¹¹	?	?	✗	✗	?	✓	✗
Hong 2021 (comparison) ¹¹	✗	✗	✗	✗	N/A	N/A	N/A
Laugwitz 2024 ²⁷	✓	✗	✓	✗	✓	✓	✓

✓ Low Risk ✗ High Risk ? Unclear Risk

N/A: not applicable; QUADAS: Quality Assessment of Diagnostic Accuracy Studies; QUADAS-C: Quality Assessment of Diagnostic Accuracy Studies-comparative

Discussion of findings

The available evidence to inform research question 1 ‘*What is the accuracy of single test and 2-tier NBS screening strategies for MLD, using DBS samples?*’ was sparse. All three publications included in this evidence summary reported early-stage studies which aimed to assess the feasibility of implementing NBS screening for MLD and all three studies were rated as having high risk of bias with respect to evaluating the accuracy of NBS screening algorithms for MLD; the key issues were in relation to the ‘flow and timing’ domain, most importantly, given that application of the diagnostic reference standard or long-term follow-up of all screen-negative babies is unlikely to be considered practicable, no publication reported a standardised approach to identifying and recording any cases missed by screening.^{11, 27, 28}

No studies were identified which reported experience from implemented screening programmes. Our supplementary searching for implemented NBS screening programmes for MLD identified a news article, from Oslo University Hospital, reporting that: ‘*In January 2025, Norway became the first country in the world to start national screening for metachromatic leukodystrophy (MLD).*’⁴¹ However, we were not able to identify any further published details about the new Norwegian screening programme.

The 2023 UKNSC evidence map⁷ noted that Hong et al. (2021)¹¹ evaluated a 2-tier screening strategy for MLD, followed by genetic confirmatory testing for clinically relevant *ARSA* variants, which was designed to have 100% sensitivity and was subsequently found to have almost 100% specificity. The 2023 UKNSC evidence map stated that further work to evaluate all available screening strategies would be justified.⁷

Hong et al. (2021) is also included in this evidence summary and reports the design and initial evaluation of a 2-tier NBS screening strategy for MLD. The 1st tier of the screening algorithm comprised the quantification of C16:0-sulfatide in DBS UPLC-MS/MS¹¹ and the 2nd tier comprised an enzymatic activity assay, using a tandem mass spectrometry method previously published by the authors.¹⁰ It was reported that the 1st tier screening cut-off required to achieve 100% sensitivity was $\geq 0.17 \mu\text{mol/L}$ and the 2nd tier screening cut-off was set at *ARSA* enzymatic

activity <20% of the mean.¹¹ The remaining two studies included in this evidence summary, a large (n=109,259) pilot study conducted in Germany²⁷ and a small (n=3,687) ‘pre-pilot’ study conducted in the UK,²⁸ both evaluated NBS screening algorithms for MLD which were based on the methods described in Hong et al. (2021).¹¹ No studies evaluating single test screening strategies were identified.

It is important to note that no study included in this evidence summary reported either confirmatory genetic testing of screen negative DBS or any method (e.g. records review or surveillance) designed to identify cases of MLD that may have been missed by screening (FN). Hence all reported or calculated estimates of the performance of NBS screening algorithms for MLD are uncertain and speculative in nature, since they assume that no cases of MLD were missed. The possibility of missed cases is highlighted by the experience of the UK pre-pilot study.²⁸

Hong et al. (2021) reported the identification of one case of MLD¹¹ and Laugwitz et al. (2024) reported the identification of three cases of MLD.²⁷ Although the first 2 tiers of the screening algorithm evaluated in Laugwitz et al. (2024) correspond to the 2-tier algorithm described in Hong et al. (2021), estimates of FPR for the 2-tier screening algorithm, derived from these two studies, were highly variable at 0.0037% and 0.016% respectively. Assuming that no cases of MLD were missed, predictive values can be calculated from the data reported in these two studies; the calculated PPVs, 50% (95% CI: 12.35 to 87.65%)¹¹ and 15% (95% CI: 9.89 to 22.11%)²⁷ are also indicative of a high level of uncertainty about the performance of this algorithm and highlight the need for further research and evaluation to increase the certainty in these estimates. The use of 3rd tier targeted genetic testing of the DBS in Laugwitz et al. (2024) illustrates the potential of this type of testing to identify screen-negative heterozygous carriers. However, depending upon how it is used to select patients for confirmatory testing, use of 3rd tier targeted genetic testing could also result in patients with novel pathogenic variants being missed by screening (false negatives).

As would be expected from the sample size (n=3,687) and the incidence of screen-detected MLD in Hong et al. (2021) and Laugwitz et al. (2024), the UK ‘pre-pilot’ study, Wu et al. (2024) did not identify any cases of MLD.²⁸ However, a new case of late infantile MLD was detected during the process of establishing reference ranges for C16:0-S levels and ARSA enzymatic activity. The DBS from this newborn had a C16:0-S level of 0.15 µmol/L and ARSA enzymatic activity of 4% of mean,²⁸ suggesting that a lower C16:0-S cut-off than that reported in Hong et al. (2021) may be needed to achieve 100% sensitivity.

Summary of findings relevant to Criteria 4 and 5

Assessment of a criterion relies on the reviewer’s consideration of the quantity, quality, applicability, and consistency of the evidence. Whilst the evidence which is available to inform assessment of UK NSC criteria 4 and 5 is promising, there is currently not sufficient evidence to conclude these criteria are met, which would require the reviewers to be confident that further research is unlikely to change the outcome. The current volume of evidence to inform this review question remains low, with publications available from only 2 pilot studies and a pre-pilot study. Although there is consistency across studies on the cut-off value for C16:0-sulfatide, there is some between-study variation in other parts of the screening and confirmatory pathway.

The limited volume of evidence currently available indicates that Criterion 4, ‘*There should be a simple, safe, precise and validated screening test.*’ is not currently met. This conclusion is based on the lack of data to inform estimates of the accuracy of evaluated screening algorithms

and the high variability in estimates of PPV calculated from two studies, for the same 2-tier screening algorithm. High uncertainty in the estimates of PPV is driven, in part, by the small number of cases reported in the studies to date. In addition, there are considerations around the practicality of the test for UK screening labs, for example the feasibility of performing 3 tiers of screening/genetic confirmatory testing on a single bloodspot.

Findings from the small UK 'pre-pilot' study included in this evidence summary indicate that Criterion 5 '*The distribution of test values in the target population should be known and a suitable cut-off level defined and agreed*' is not currently met. In particular, the incidental identification of a new case of late infantile MLD, during the validation phase of this study suggests that more research and evaluation may be required to specify the cut off for the first tier of testing; DBS from this newborn had a C16:0-S level of 0.15 $\mu\text{mol/L}$, which is below the ≥ 0.17 $\mu\text{mol/L}$ cut-off used in the 2-tier algorithm evaluated by all three of the studies included in this evidence summary and which has been reported as the cut-off required to achieve 100% sensitivity.¹¹ While detection rates do appear to be high in the two studies which detected cases of MLD, it is not possible to know if this could be overestimated due to a lack of confirmatory testing or follow up in screen negative babies.

The move to larger scale studies in screening populations is relatively recent and represents a step forward in the evidence around screening for MLD. There are several ongoing pilot and implementation studies in other countries; future publication of data from such trials has the potential to provide further evidence to inform UK NSC criteria 4 and 5. Norway has recently implemented screening for MLD, and the publication of findings from the Norwegian programme would further expand the evidence base. Further studies of the performance of screening algorithms in UK newborns would also provide valuable information to develop and expand the evidence base.

Criterion 9 - Efficacy of treatment in the pre-symptomatic phase

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.

Question 2 - Does early initiation of treatment following screening lead to improved outcomes for MLD compared to initiation of treatment following clinical presentation?

The 2023 UKNSC evidence map,⁷ identified 19 studies,^{29, 40, 47-63} which were relevant to the question:

What is the volume and type of evidence available on the benefits and/or harms of interventions in presymptomatic/asymptomatic children with MLD identified through screening?

These studies included 14 interventional studies,^{29, 40, 47-57, 61} four cohort studies,^{59, 60, 62, 63} and one case-control study.⁵⁸ Where reported, the sample size ranged from 2 to 12 presymptomatic patients. All studies reported on the benefit and/or harms of treatment in presymptomatic children with MLD. Importantly, none of these studies reported on patient populations identified

through newborn screening or compared outcomes between patients who were identified and treated before the development of symptoms and those treated following symptomatic presentation.

The 2023 UKNSC evidence map,⁷ noted that Libmeldy® was the most commonly evaluated intervention (14 publications)^{29, 40, 47-57, 60} and was found to be effective, safe and well tolerated for the treatment of patients with pre-symptomatic MLD. However, it should be noted that all but two^{29, 40} of these publications were conference abstracts and all publications had authors in common, indicating a potential for overlapping populations and multiple reporting of results.

The 2023 UKNSC evidence map concluded that: 'At present, there is sufficient evidence on the effects of treatments for MLD to recommend further work on this question.'⁷

What is added by this evidence summary

This evidence summary provides a summary of the published studies available to inform question 2, which includes the two full publications, Fumagalli et al. (2022)²⁹ and Sessa et al. (2016)⁴⁰ identified by the 2023 UKNSC evidence map,⁷ and one additional study identified in the systematic review by Groeschel et al. (2016).³⁰

We did not identify any studies which compared the efficacy of treatments for MLD in early (screening or cascade testing) versus late (symptomatic presentation) detection.

Description of new evidence in relation to previous evidence reviews

The searching and title and abstract screening stages of the evidence summary were conducted as a single process, with consideration of all three research questions. Appendix 2 provides an overall PRISMA flow chart for this evidence summary and details of studies included and excluded after full-text screening.

Following full-text screening, there were three publications that met the inclusion criteria specified for research question 2.^{29, 30, 40} Two of these publications, Fumagalli et al. (2022)²⁹ and Sessa et al. (2016),⁴⁰ reported outcomes from the same study (NCT01560182), a phase I/II clinical trial of the gene therapy Libmeldy® for the treatment of patients with presymptomatic or early-symptomatic MLD, partially funded by the manufacturer of ARSA-cel (Orchard Therapeutics). Both of these publications also reported the results of analyses comparing outcomes in treated patients to an untreated natural history (NH) cohort of patients with early-onset MLD, matched by age and disease subtype, and both publications were included in the 2023 UKNSC evidence map⁷ and the primary publication reporting the most recent and largest data set was Fumagalli et al. (2022).²⁹ It should be noted that an unpublished comparison between patients from NCT01560182 and patients from a NH cohort, matched by age and disease subtype (late infantile or early juvenile), formed the basis of the clinical effectiveness evidence (summarised and critiqued in the External Assessment Group [EAG] report⁶⁴) that informed NICE guidance HST18: 'Atidarsagene autotemcel for treating metachromatic leukodystrophy.'⁵ Fumagalli et al. (2022), included in this evidence summary, is a subsequent publication which appears to report results for the same comparison as the EAG report (same data sources and methods are reported). The remaining publication, identified by this evidence summary, was a retrospective observational study that compared long-term the outcomes of patients with juvenile MLD, who underwent allogeneic HSCT with a control cohort who did not undergo allogeneic HSCT.³⁰ Study details, participant characteristics and details of the treatments evaluated are provided in Tables 7, 8 and 9, respectively.

None of the three publications included in this evidence summary compared the efficacy of treatments for MLD in screen- or cascade test-detected MLD versus symptomatic MLD. All three publications were therefore included on the basis of the secondary eligibility criteria described in the footnotes to Table 2. All three publications reported retrospective analyses, comparing the treatment outcomes of patients with pre-symptomatic/early symptomatic MLD to the outcomes of untreated patients (NH cohorts), to estimate treatment effects.^{29, 30, 40} The results of these comparisons are summarised in Tables 10 and 11. In addition, two publications reported some assessment of correlation between time to treatment and outcome.^{30, 40}

Following consultation, an additional publication, Fumagalli et al. (2025)⁶⁵, was identified which would have met the inclusion criteria for this evidence review had it been published prior to the search end date of 03 Feb, 2025. This study was not included in the original review but was evaluated using the same methodology and is described in the ‘post-consultation addendum’ below. Although not strictly an included study, Fumagalli et al. (2025)⁶⁵ has been added to Tables 7-12 for completeness.

Fumagalli et al. (2022)²⁹ and Sessa et al. (2016)⁴⁰ (NCT01560182)

Fumagalli et al. (2022) reported outcomes for paediatric patients treated with Libmeldy® in a prospective, non-randomised, phase 1/2 clinical trial (NCT01560182) and expanded-access frameworks; gross motor function was the primary outcome. Treated patients comprised 16 patients with late infantile MLD (15 pre-symptomatic and one early symptomatic) and 13 patients with early juvenile MLD (five pre-symptomatic and eight early symptomatic).²⁹ At the time of publication, 26/29 treated patients were alive, with a median follow-up of 3.16 years (range 0.64 to 7.51 years). All deaths occurred in patients with early symptomatic early juvenile MLD; there were two deaths due to disease progression and one death due to ischaemic stroke following an infectious event 13.6 months after treatment (deemed to be unrelated to treatment).²⁹ With respect to the analyses comparing outcomes between treated and untreated patients, NH patients selected for the matched analyses were classified as having late infantile or early juvenile MLD, and had a study visit at which their age fitted within the age range for treated patients with the same disease sub-type at the time of analysis. All NH patients were symptomatic at the time of enrolment, but retrospective collection of data for the period prior to enrolment enabled aged-matched analyses.²⁹

Comparisons of the Gross Motor Function Measure (GMFM) score, between treated and NH patients were presented by MLD disease sub-type (late infantile and early juvenile), and further subgroup analyses were presented for pre-symptomatic and early symptomatic patients with juvenile MLD.²⁹ The results indicated that treatment was associated with improved GMFM, relative to NH, for all patient groups, at 3 years after treatment (the longest reported follow-up point). Of note, the adjusted mean difference (MD) in GMFM at 3 years between the treated and NH groups appeared greater in the pre-symptomatic early juvenile MLD subgroup than in the early symptomatic MLD subgroup (74.9 [95% CI: 50.8 to 99.1] versus 43.9 [95% CI: 9.2 to 78.5]); for the early symptomatic early juvenile MLD sub-group the treatment effect did not reach statistical significance at the 5% level (see Table 11).²⁹ The adjusted MD in GMFM at 3 years between the treated and NH groups was 71.5 (95% CI: 50.3 to 92.7), in the late infantile MLD group (where 15/16 treated patients were pre-symptomatic).²⁹ Insufficient data were reported to allow a direct comparison of outcomes between, e.g. the treated pre-symptomatic early juvenile and the treated early symptomatic early juvenile groups. Fumagalli et al. (2022) also reported comparisons of severe motor impairment-free survival between treated and NH patients; reporting of treatment effect for this outcome was in the form of Kaplan-Meier (KM) plots and *p*-values, and the treatment effect was statistically significant in the late infantile MLD group only (see Table

10). At 4.5 years of age, 92% (95% CI: 57 to 99%) of patients with late infantile MLD, who were treated with Libmeldy®, were alive and free from severe motor impairment (Gross Motor Function Classification [GMFC] level ≤ 4 , able to sit, crawl or roll, or better), compared to zero patients with late infantile MLD in the NH group.²⁹

Comparisons of brain MRI severity score, between treated and NH patients, were presented by MLD disease sub-type (late infantile and early juvenile). The results indicated that treatment was associated with improved brain MRI score, relative to NH, for both disease subtypes, at 3 years after treatment (see Table 11).²⁹

A comparison of nerve conduction velocity (NCV) index, between treated and NH patients, was presented for the late infantile MLD group only and the results indicated that treatment was associated with improved NCV, relative to NH, at 3 years after treatment (see Table 11).²⁹

Sessa et al. (2016) reported outcomes for a sub-set of nine patients from NCT01560182 (six patients with late infantile MLD, two patients with early juvenile MLD and one patient whose disease subtype could not be definitively classified).⁴⁰ This publication noted that the extent of treatment benefit appeared to be influenced by the interval between HSC-GT and the expected time of disease onset; change in GMFM score, from baseline to last follow-up positively correlated with the time interval between HSC-GT administration and expected or actual disease onset (Spearman $r=0.8034$, $p=0.0138$).⁴⁰

Groeschel et al. (2016)³⁰

Groeschel et al. (2016) reported a retrospective study comparing long-term outcomes of patients with juvenile MLD who underwent allogeneic HSCT in one of three German centres and who were followed up for at least 2 years with control patients with juvenile MLD, from a NH study within the German leukodystrophy network Leukonet, who did not undergo transplant.³⁰ Transplanted patients were born between 1975 and 2009 and control patients were born between 1967 and 2007. All transplanted patients had pre-symptomatic or early symptomatic juvenile MLD at the time of transplant. No details of the symptom stage of control patients at diagnosis/initial assessment were reported and the analysis did not appear to have included any attempt to match treated patients and controls with respect to prognostic factors. All-cause mortality in transplanted patients, at a median follow-up of 7.5 years (range 3 to 19.7 years), was 6/24 (25%). This was similar to the all-cause mortality rate for control patients, 11/41 (27%), (follow-up duration not reported); whilst MLD progression related mortality appeared lower in transplanted patients, 2/24 (8.3%), than in controls, where all deaths were MLD progression-related, any potential effect on overall mortality was lost due to four transplant-related deaths.³⁰

Comparisons of MLD progression reported in Groeschel et al. (2016) indicated that HSCT was associated with reductions in both the rate of progression to severe motor impairment (GMFC level 5, loss of locomotion and sitting without support) and the rate of language loss, at 10 years after disease onset (see Table 10). Calculated effect estimates and reported p values indicated borderline statistical significance, reflecting the small sample size.³⁰

Transplant also appeared to be associated with improved brain MRI severity score, as measured at the last imaging session after HSCT, compared to control patients, however, no time point was reported for the measurement of comparator brain MRI severity score in control patients.³⁰

Groeschel et al. (2016) also included results from an exploratory multivariable analysis with relevant independent variables (including symptom status at HSCT).³⁰ Predictors of prognostic parameters for stable versus progressive disease after HSCT were: patients who underwent HSCT at GMFC-MLD levels 0 and 1 ($p=0.02$); patients who underwent HSCT with an IQ ≥ 85 ($p=0.02$); age at onset older than 4 years ($p=0.01$). MRI severity score >17 was associated with disease progression ($p=0.03$). Citing the retrospective design and small sample size of their study, the study authors emphasised that these analyses were explorative rather than confirmatory in nature and that all p values were, therefore, considered descriptive.³⁰

Methodological quality of studies

Fumagalli et al. (2022) and Groeschel et al. (2016) were assessed using the ROBINS-I tool for assessing risk of bias in non-randomised studies of interventions;³⁷ no assessment was conducted for Sessa et al. (2016) as this is a secondary publication relating to the same study as Fumagalli et al. (2022), (NCT01560182). Fumagalli et al. (2022) and Groeschel et al. (2016) were rated as being at serious and critical risk of bias, respectively, with the key area of concern being inadequate or absent consideration of potential confounding. A summary of the results of the ROBINS-I assessments is provided in Table 12 and full assessment are provided in Appendix 3.

Discussion of findings

The available evidence to inform research question 2 ‘Does early initiation of treatment following screening lead to improved outcomes for MLD compared to initiation of treatment following clinical presentation?’ was sparse and of low methodological quality. This reflects the significant challenges associated with conducting research in a rare disease setting.

None of the three publications included in this evidence summary compared the efficacy of treatments for MLD in in screen- or cascade test-detected MLD versus symptomatic MLD.

All three of the publications included in this evidence summary provide some limited information about the effects of treatment in patients with pre-symptomatic/early symptomatic MLD, compared to untreated patients (NH cohorts).^{29, 30, 40}

Fumagalli et al. (2022) provides some evidence that the gene therapy Libmeldy®, recommended in NICE guidance HST18,⁵ may be effective in improving gross motor function for patients with pre-symptomatic/early symptomatic late infantile or early juvenile MLD, relative to an untreated NH cohort.²⁹ All comparisons reported in Fumagalli et al (2022) were for treated vs. untreated patients; no comparative effects were reported for treatment in pre-symptomatic vs. early symptomatic patients. Observational comparison of the values for adjusted MD in GMFM reported for disease type and symptom status subgroups appears to indicate larger point estimates of effect size in the subgroup of patients with early juvenile MLD who were treated before the onset of symptoms compared to the treated group with early symptomatic early juvenile MLD. However, the small sample size (reduced further in subgroup analyses) mean that this observation should be considered hypothesis generating and not evidential.

The EAG report for NICE guidance HST18⁶⁴ redacts most of the clinical effectiveness results.. However, it should be noted that the key areas of concern raised by the EAG are also applicable to the version of the analysis published in Fumagalli et al. (2022).²⁹ The EAG considered that age or predicted age at disease onset should be an important prognostic characteristic to consider in any MLD patient matching exercise.⁶⁴ The EAG also noted limitations in the

reporting of results and of baseline data for both the treated and NH cohorts,⁶⁴ limitations which were also present in Fumagalli et al. (2022).²⁹ The EAG noted that the evidence for the effectiveness of patients treated with Libmeldy® was most substantial in those with late infantile MLD;⁶⁴ this observation is consistent with the results published in Fumagalli et al. (2022).²⁹ The EAG also noted that some patients (particularly in the early symptomatic early juvenile MLD group) showed a decline in motor function and it is unclear whether these patients will stabilise with impairment or continue to decline, and that treated patients with early juvenile MLD still experience peripheral neuropathy (NCV index scores were as bad or worse than those in NH patients),⁶⁴ these data were not included in Fumagalli et al. (2022).²⁹

The only study to evaluate a non-gene therapy treatment, Groeschel et al. (2016), used a retrospective study design and an indirect comparison to evaluate the effectiveness of HSCT in patients with juvenile MLD, relative to untreated control patients.³⁰ The results of this study indicated that transplant was associated with reductions in both the rate of progression to severe motor and the rate of language loss, for surviving patients at 10 years after disease onset. In addition, exploratory analysis indicated that transplant before or in the very early stages of symptom onset (GMFC-MLD level 0 or 1) may be predictive of stable disease following HSCT. However, it should be noted that the high rate of transplant-related mortality (16.7%) resulted in similar rates of all-cause mortality between the transplanted (25%) and control (27%) groups, raising questions about the overall effectiveness of HSCT as a potential treatment. As with Fumagalli et al. (2022), there is uncertainty around the findings of this study due to the small sample size and weakness of the study design; Groeschel et al. (2016) was considered to be of lower methodological quality because no attempt to match treated patients and controls with respect to prognostic factors. There is further uncertainty, with respect to the applicability of the study to current practice, in that all transplants occurred between 1991 and 2012 and transplant outcomes (e.g. transplant-related mortality) may not be representative of those achievable by current best practice.

Post-consultation addendum

This addendum provides a summary of the additional data presented in Fumagalli et al. (2025),⁶⁵ which was published after completion of the rapid evidence review. As with Fumagalli et al. (2022),²⁹ Fumagalli (2025) compared the outcomes of patients with pre-symptomatic/early symptomatic MLD, who were treated with Libmeldy® to the outcomes of untreated patients from a NH cohort.

Fumagalli et al. (2025),⁶⁵ reported long-term follow-up, median 6.76 years (range: 0.64 to 12.19 years), for the 29 patients (20 treated in N C T01560182 and 9 treated through expanded access programmes) who were previously included in Fumagalli et al. (2022).²⁹ Fumagalli et al. (2025)⁶⁵ included results for an additional 10 treated patients from a separate, single arm, open label study (NCT03392987) which evaluated the use of cryopreserved autologous CD34+ Cells transduced with lentiviral vector containing human ARSA CDNA (Libmeldy®). Two patients, both treated in N C T01560182, were excluded from the integrated efficacy analysis presented in Fumagalli et al. (2025).⁶⁵ These patients had evidence of rapid disease progression during the period shortly before treatment and were adjudicated by the IRC as being cases of symptomatic LI MLD (1 patient) and progressively symptomatic EJ MLD (1 patient); they would not, therefore, meet current eligibility criteria for treatment. Both patients were included in the analyses presented in Fumagalli et al. (2022).²⁹ Individual plots of GMFC-MLD level versus age were presented (supplementary materials) for both these patients compared to an untreated sibling and/or matched patients from the NH cohort; in both cases the deterioration in GMFC over time appeared similar to that of the untreated comparator.⁶⁵

Study details and results for Fumagalli et al. (2025)⁶⁵ have been added to Tables 7 to 11, below. In addition to extended follow-up, Fumagalli et al. (2025),⁶⁵ reported results for additional outcome measures (motor impairment-free survival, cognitive impairment-free survival, loss of language and progression to severe motor function impairment or death). Results were reported, for all outcome measures, for the pre-symptomatic and early symptomatic subgroups of patients treated with Libmeldy®, however, no formal statistical comparisons between these groups were reported and the numbers of patients were very small for some subgroup-outcome-timepoint combinations. Small sample sizes and wide and overlapping confidence intervals make meaningful observational comparisons difficult. Of note, the apparent difference in the point estimate of treatment effect (adjusted mean difference in GMFM) between the pre-symptomatic and early symptomatic early juvenile MLD subgroups, observed at 3 years,²⁹ appeared to be maintained at 5 years (Table 11).⁶⁵ Point estimates also indicated that rates of cognitive impairment-free survival may be higher and rates of language loss lower, at age 10 years, in treated patients with pre-symptomatic early juvenile MLD than in treated patients with early symptomatic MLD.⁶⁵ Rates of progression to severe motor function impairment or death were similar in the treated pre-symptomatic early juvenile (25%) and treated early symptomatic early juvenile (22%) groups, at 5 years post-treatment (Table 10).⁶⁵

The study authors also noted that all 18 treated patients with late infantile MLD were alive at the time of analysis and had a median age of 7.4 years (range: 3.2 to 13.4); 19/28 untreated patients with late infantile MLD had died and their median age at death was 5.9 years (range: 3.5 to 13.4). Overall survival was stated to be similar among treated patients with pre-symptomatic early juvenile MLD and those with early symptomatic early juvenile MLD (1 patient with pre-symptomatic early-juvenile MLD died from an ischemic cerebral infarction at 2.1 years of age, and 2 patients with early symptomatic early-juvenile MLD died due to disease progression).⁶⁵

A separate ROBINS-I assessment has been conducted for Fumagalli et al. (2025);⁶⁵ this has been added to Appendix 3 and is summarised in Table 12. There were some differences in the rating of individual domains of ROBINS-I for Fumagalli et al. (2025)⁶⁵ compared to Fumagalli et al. (2022),²⁹ however, the overall risk of bias rating remained the same.

In summary, observational comparisons of the extended follow-up and additional outcomes data provided by Fumagalli et al. (2025)⁶⁵ are consistent with observations derived from Fumagalli et al. (2022).²⁹ In general, outcomes following treatment with Libmeldy® appear better for patients with pre-symptomatic early juvenile MLD than for those with early symptomatic early juvenile MLD. However, the observational nature of these comparisons and the very small sample sizes involved mean that they should be interpreted with caution. Individual patient data for two treated patients who were excluded from the analyses (1 with symptomatic late infantile MLD and 1 with progressive early juvenile MLD), showing deterioration in GMFC over time, indicated that treatment with Libmeldy® was ineffective for these two patients; this lack of effectiveness provides some additional support for a difference in the effectiveness of treatment with Libmeldy® between pre-symptomatic and symptomatic patients.⁶⁵

There remains no direct evidence that treatment following identification of patients with MLD through screening or cascade testing results in improved outcomes relative to treatment following symptomatic presentation. However, it should be noted that, given the licensed indication and NICE recommendation for Libmeldy®, the challenges of generating direct evidence should not be underestimated.

Summary of Findings Relevant to Criterion 9

The limited evidence available indicates that 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.’* is currently not met.

This review question does not seek to answer whether Libmeldy is effective as a treatment for late infantile and early juvenile MLD, as this has already been considered by NICE. This review question is instead examining whether there is evidence that earlier treatment of individuals with MLD identified via screening; either with Libmeldy, allogenic HSCT, or any other treatment; leads to improved outcomes for the screened individual compared with usual care. This information is essential for the UK NSC to determine whether the potential harms imposed by screening the whole newborn population (for example from potential false positive results, false negative results, uncertain findings, overdiagnosis, overtreatment), which may affect any screened individual, are outweighed by the benefits of earlier detection in individuals with MLD. As noted in the footnote of table 2, if no studies were identified which directly compared the efficacy of treatments for MLD detected early (screening or cascade testing) versus late (symptomatic presentation), then publications were included which compared the treatment outcomes of patients with pre-symptomatic/early symptomatic MLD to the outcomes of untreated patients (natural history cohort) to estimate treatment effects.

Even with the inclusion of indirect evidence, and the additional consideration of Fumagalli et al (2025)⁶⁵, the available evidence base is small in volume and has methodological limitations. This was also acknowledged by the NICE evidence review group (ERG) on pages 639 and 700 of the HST18 evaluation consultation committee papers⁵; *“There is currently, no evidence to inform the effect of new born screening on the effectiveness of OTL-200 [Libmeldy]”* and *“While the ERG recognises that the limitations of the available evidence, the ERG is broadly in agreement with the company that the introduction of new-born screening would positively impact on the effectiveness of OTL-200. [...] Whether the introduction of new-born screening would lead to all patients achieving a full response as modelled in the company’s scenario is a matter of speculation, but the ERG considers it reasonable to expect that a greater proportion of patients who are currently classified as stabilised partial responders would achieve full response.”*

There is some very weak, indirect evidence to indicate that the effects of gene therapy treatment (Libmeldy®) on gross motor function, relative to untreated patients, may be greater for patients with early juvenile MLD when treatment is received before symptoms develop. This evidence was derived from subgroup analyses of patients with pre-symptomatic and early symptomatic early juvenile MLD reported in a study with substantial methodological limitations, which was partially funded by Orchard Therapeutics (the manufacturer of Libmeldy®); no comparisons of treatment effect between pre-symptomatic and symptomatic patients, were possible for late infantile MLD since all but one of the patients with late infantile MLD, who were treated with Libmeldy®, received treatment before symptoms developed.

There is evidence that treatment of late infantile and early juvenile MLD with Libmeldy® before symptoms develop is effective compared to no treatment. Importantly, there is currently no

direct evidence evaluating the way in which outcomes from the treatment are affected by its delivery within a newborn screening setting compared to no screening. The challenges of generating this kind of evidence should not be underestimated. However, approaches to evaluation are available; for example, strategies similar to the UK in-service evaluation for severe combined immunodeficiency⁶⁶ and the German non-randomised controlled trial for spinal muscular atrophy (SMA)⁴⁵ could provide important evidence on the effectiveness of newborn screening while increasing the number of cases studied.

Table 7: Details of included treatment studies

Publication (Study)	Country	Study design	Participant selection criteria	Timeframe	Outcomes extracted
Fumagalli 2025 ⁶⁵ (NCT01560182 and NCT03392987)		Observational: retrospective integrated analysis of results from two prospective, non-randomised, phase 1/2 clinical studies and expanded access networks (patients treated with ARSA-cel gene therapy) compared with an untreated NH cohort	Inclusion criteria: As for Fumagalli 2022 ²⁹ Exclusion criteria: Not specified Comparator group: As for Fumagalli 2022 ²⁹	2004 to 2022	All-cause mortality;
Fumagalli 2022 ²⁹ (NCT01560182)	Italy	Observational: retrospective integrated analysis of results from a prospective, non-randomised, phase 1/2 clinical study and expanded access networks (patients treated with ARSA-cel gene therapy) compared with an untreated NH cohort	Inclusion criteria: patients with a molecular and biochemical diagnosis of MLD of either pre-symptomatic late-infantile, or pre-symptomatic or early symptomatic early-juvenile variants. Late-infantile MLD was diagnosed when the predicted age at symptom onset was 30 month or less based on an index sibling, early-juvenile MLD was diagnosed when the predicted age at onset based on the index sibling or patient's own age at disease onset was 30 months to 6 years. Early symptomatic status was initially defined as presence of symptoms for <6 months; the definition was expanded (January 2014) to allow treatment of patients with IQ ≥70 and the ability to walk 10 or more steps independently, and to exclude severely impaired patients or those entering rapid disease progression. Exclusion criteria: positive test for HIV or hepatitis C or B; affected by neoplastic diseases; cytogenetic alterations typical of myelodysplastic syndrome or acute myelogenous	2004 to 2017	All-cause mortality; GM FM score; severe motor impairment-free ^a survival; brain MRI severity score; NCV index

			<p>leukaemia; end-organ function or other severe disease which, as judged by the investigator, would make the patient inappropriate for study entry; enrolment in other trials; had undergone allogenic HSCT in the previous 6 months; had undergone allogenic HSCT with evidence of residual cells of donor origin.</p> <p>Comparator group: patients selected from a NH study (SR-Tiget, no reference reported) were matched by age and MLD disease subtype (late infantile or early juvenile)</p>		
Sessa 2016 ⁴⁰ (NCT01560182)	Italy	<i>Post-hoc</i> analysis of data from an ongoing, non-randomised, open label, single arm phase 1/2 trial	<p>Inclusion criteria: patients with a molecular and biochemical diagnosis of MLD of either pre-symptomatic late-infantile, or pre-symptomatic or early symptomatic early-juvenile variants. Late-infantile MLD was diagnosed when the predicted age at symptom onset was 30 months or less based on an index sibling, early-juvenile MLD was diagnosed when the predicted age at onset based on the index sibling or patient's own age at disease onset was 30 months to 6 years. Early symptomatic status was defined as presence of symptoms for <6 months.</p> <p>Exclusion criteria: positive test for HIV or hepatitis C or B; affected by neoplastic diseases; cytogenetic alterations typical of myelodysplastic syndrome or acute myelogenous leukaemia; end-organ function or other severe disease which, as judged by the investigator, would make the patient inappropriate for study entry; enrolment in other trials; had undergone allogenic HSCT in the previous 6 months; had</p>	2010 to 2013	Correlation analysis only (no additional outcomes)

undergone allogenic HSCT with evidence of residual cells of donor origin.

Comparator group: historical cohort of 21 patients with late-infantile MLD and nine patients early juvenile MLD, from the LDM1 study,⁶⁷ who had not received treatment.

Groeschel 2016 ³⁰	Germany	Retrospective 'case-control' analysis comparing long-term outcomes of patients with juvenile MLD who underwent allogenic HSCT with control patients who did not undergo transplant	<p>Inclusion criteria: patients who were followed up for at least 2 years after HSCT in one of three German centres.</p> <p>Exclusion criteria: none reported</p> <p>Comparator group: non-transplanted controls from a natural history study within the German leukodystrophy network Leukonet</p>	<p>Treated patients: born between 1975 and 2009</p> <p>Control patients: born between 1967 and 2007</p>	All-cause mortality; progression to GMFC-MLD level 5 ^b ; language loss; brain MRI severity score
------------------------------	---------	--	--	---	---

^a Surviving and maintaining GMFC level ≤ 4 (able to sit, crawl or roll, or better).

^b Loss of locomotion and sitting without support (severe motor impairment).

ARSA: Arylsulfatase A; GMFC-MLD: Gross Motor Function Classification in MLD; GMFM: Gross Motor Function Measure; HIV: human immunodeficiency virus; HSCT: hematopoietic stem cell transplantation; IQ: intelligence quotient; MLD: metachromatic leukodystrophy; MRI: magnetic resonance imaging; NCV: nerve conduction velocity; NH: natural history

Table 8: Participant characteristics for treatment studies

Publication (Study)	Group (N)	Sex, n (%)	Ethnicity	Disease stage	Age at treatment or age at initial assessment for untreated cohorts, mean (SD) or median (range)
Fumagalli 2025 ⁶⁵ (NCT01560182 and NCT03392987)	PSLIMLD (18) ^a	M, 13 (72) F, 5 (28)	Hispanic/Latino, 1 (6%) White/Caucasian, 13 (72%) Arabic/N African, 3 (17%) Asian, 2 (12%)	N/A	10.3 (7.6 to 17.7) months

	PSEJMLD (8)	M, 6 (75) F, 2 (25)	African American/African, 0 (0%) Hispanic/Latino, 0 (0%) White/Caucasian, 6 (75%) Arabic/N African, 1 (13%) Asian, 0 (0%)	N/A	16.1 (11.3 to 48.9) months
	ESEJMLD (11) ^b	M, 6 (55) F, 5 (45)	African American/African, 1 (13%) Hispanic/Latino, 0 (0%) White/Caucasian, 11 (100%) Arabic/N African, 0 (0%) Asian, 0 (0%)	N/A	69.0 (30.5 to 139.7) months
	NHLIMLD (26)	M, 12 (46) F, 14 (54)	African American/African, 0 (0%) Hispanic/Latino, 2 (8%) White/Caucasian, 23 (88%) Arabic/N African, 3 (12%) Asian, 0 (0%)	NR	18.8 (14.5 to 27.9) months
	NHEJMLD (17)	M, 8 (47) F, 9 (53)	African American/African, 0 (0%) Hispanic/Latino, 0 (0%) White/Caucasian, 16 (94%) Arabic/N African, 1 (6%) Asian, 0 (0%)	NR	52.6 (19.2 to 74.1) months
Fumagalli 2022 ²⁹ (NCT01560182)	ITTLIMLD (16)	M, 10 (62) F, 6 (38)	African American/African, 0 (0%) South-East Asian, 1 (6%) Arabic/N African, 4 (25%) White/Caucasian, 11 (69%)	pre-symptomatic, 15 (94%) early symptomatic, 1 (6%)	12.81 (4.3) months
	ITTEJMLD (13)	M, 6 (46) F, 7 (54)	South-East Asian, 0 (0%) Arabic/N African, 0 (0%) White/Caucasian, 13 (100%)	pre-symptomatic, 5 (38%) early symptomatic, 8 (62%)	65.86 (33.4) months
	NHLIMLD (19)	M, 8 (42) F, 11 (58)	South-East Asian heritage, 0; Arabic/North African heritage,	pre-symptomatic, 0 (0) early symptomatic, 19 (100%)	20.64 (4.7) months

			3 (16%); White/Caucasian heritage, 16 (84%)		
	NHEJMLD (12)	M, 5 (42) F, 7 (58)	South-East Asian, 0 (0%) Arabic/N African, 0 (0%) White/Caucasian, 12 (100%)	pre-symptomatic, 0 (0) early symptomatic, 12 (100%)	51.98 (19.2) months
Sessa 2016 ⁴⁰ (NCT01560182)	LI MLD (6); LIVEJ MLD (1); EJMLD (2)	NR	Caucasian, 5 (56) Arabian, 3 (33) Hispanic, 1 (11)	pre-symptomatic, 5 (56%) early symptomatic, 4 (44%)	16 (7 to 59) months
Groeschel 2016 ³⁰	Treated juvenile MLD (24)	M, 11 (46) F, 13 (54)	NR	pre-symptomatic, 10 (42) FMS only, 1 (4) CD only, 3 (12.5) GMFC-MLD1, 5 (21) GMFC-MLD2, 3 (12.5) GMFC-MLD3, 2 (8)	7 (1.5 to 18.2) years
	Control juvenile MLD (41)	M, 22 (54) F, 19 (46)	NR	NR	6.5 (2.7 to 16) years ^c

^aOne patient, determined by the independent review committee to have symptomatic late infantile MLD, was excluded from the integrated efficacy analyses

^bOne patient, determined by the independent review committee to have progressively symptomatic early-juvenile MLD, was excluded from the integrated efficacy analyses

^cAge at disease onset.

CD: cognitive decline; EJ: early juvenile; ES: early symptomatic; F: female; FMS: fine motor symptoms; GMFC: Gross Motor Function Classification; ITT: intention-to-treat; LI: late infantile; M: male; MLD: metachromatic leukodystrophy; N: North; N/A: not applicable NH: natural history; NR: not reported; PS: presymptomatic; SD: standard deviation

Table 9: Treatment details

Publication (Study)	Gene therapy	Donor, n (%)	Graft, n (%)	Conditioning, n (%)	GvHD prophylaxis, n (%)
---------------------	--------------	--------------	--------------	---------------------	-------------------------

Fumagalli 2025 ⁶⁵ (NCT01560182 and NCT03392987)	Libmeldy®; lenti-viral vector encoding human ARSA cDNA	N/A	NR	Bus (target cumulative AUC 67,200 µg*h/L), 13 (33) Bus (target cumulative AUC 85,000 µg*h/L), 26 (67)	N/A
Fumagalli 2022 ²⁹ (NCT01560182)	Libmeldy®; lenti-viral vector encoding human ARSA cDNA	N/A	BM, 26 (90) BM + MPB, 2 (7) MPB, 1 (3)	Bus (sub myeloablative, target AUC 63.4 to 84.3 mg x h/L), 13 (45) Bus (myeloablative, target AUC 78.0 to 88.3 mg x h/L), 16 (55)	N/A
Sessa 2016 ⁴⁰ (NCT01560182)	Libmeldy®; lenti-viral vector encoding human ARSA cDNA	N/A	BM, 9 (100)	Bus (target cumulative dose 11.2 to 16.8 mg/kg), administered every 6 hours over 4 days in 14 doses (individual dose target AUC 4.8 mg x h/L)	N/A
Groeschel 2016 ³⁰	N/A	MRD, 3 (12.5) MMRD, 4 (17) MUD, 14 (58) MMUD, 3 (12.5)	BM, 17 (71) PB, 6 (25) CB, 1 (4)	Bus + CP, 1 (4) Bus + CP + ATG, 9 (37.5) Bus + CP + Thy, 2 (8) Bus + Flu + CP + Alem, 2 (8) Flu + Treo + Thio + Thy, 6 (25) Flu + Bus + Mel + ATG, 3 (12.5) Flu + Thio + Mel + OKT3®, 1 (4)	MTX + CsA, 15 (63) CsA, 7 (29) MMF, 2 (8)

Alem: alemtuzumab; ARSA: Arylsulfatase A; ATG: antithymocyte globulin; AUC: area under the curve; BM: bone marrow; Bus: busulfan; CB: cord blood; CP: cyclophosphamide; CsA: cyclosporin A; DNA: deoxyribonucleic acid; Flu: fludarabine; GvHD: graft versus host disease; Mel: melphalan; MMF: mycophenolate mofetil; MMRD: mismatched related donor; MMUD: mismatched unrelated donor; MPB: mobilised peripheral blood; MRD: matched related donor; MTX: methotrexate; MUD: matched unrelated donor; N/A: not applicable; NR: not reported; PB: peripheral blood; Thio: thiotepa; Thy: thymoglobulin; Treo: treosulfan

Table 10: Dichotomous outcomes from treatment studies

Publication (Study)	Group (N)	Number with event n/N (%) or estimated percentage with outcome (95% CI)	Follow-up, median (range)	Treatment effect
All-cause mortality				

Fumagalli 2025 ⁶⁵ (NCT01560182 and NCT03392987)	Treated PS LI MLD (18)	0/18 (0)	6.7 (2.4 to 12.2) y	-
	NH LI MLD (28)	19/28 (68) ^a	4.4 (0.6 to 18.8) y	
Fumagalli 2022 ²⁹ (NCT01560182)	Treated all MLD (29)	3/29 (10)	3.6 (0.64 to 7.51) y	-
	NH MLD (31)	NR		
Groeschel 2016 ³⁰	Treated juvenile M LD (24)	6/24 (25) ^b	7.5 (3.0 to 19.7) y	-
	Control juvenile ML D (41)	11/41 (27) ^c	NR	
Severe motor impairment-free survival^d				
Fumagalli 2025 ⁶⁵ (NCT01560182 and NCT03392987)	Treated PS LI MLD (18)	100 (95% CI: 100 to 100) ^e	at 6.0 y of age	Treated PS LI MLD versus NH LI MLD, <i>p</i> <0.001
	NH LI MLD (28)	0 (95% CI: NE)	at 6.0 y of age	
	Treated PS EJ MLD (8)	87.5 (95% CI: 38.7 to 98.1)	at 10.0 y of age	Treated PS EJ MLD versus NH LI MLD, <i>p</i> =0.04
Treated ES EJ MLD (11)	80.0 (95% CI: 40.9 to 94.6)	at 10.0 y of age		
Fumagalli 2022 ²⁹ (NCT01560182)	NH EJ MLD (21)	11.2 (95% CI: 0.9 to 36.4)	at 10.0 y of age	Treated ES EJ MLD versus NH LI MLD, <i>p</i> <0.001
	Treated LI MLD (16)	92 (95% CI: 57 to 99) ^f	at 4.5 y of age	Treated LI MLD versus NH LI MLD, <i>p</i> <0.0001
	NH LI MLD (19)	0 (0)	at 4.5 y of age	
	Treated PS EJ ML D (5)	80 (95% CI: 20 to 97)	at 8.0 y of age	Treated PS EJ MLD versus NH EJ MLD, <i>p</i> =0.173
	Treated ES EJ ML D (8)	63 (95% CI: 23 to 86)	at 8.0 y of age	
	NH EJ MLD (12)	36 (95% CI: 9 to 65)	at 8.0 y of age	Treated ES EJ MLD versus NH EJ MLD, <i>p</i> =0.031
Motor impairment-free survival^g				
Fumagalli 2025 ⁶⁵ (NCT01560182 and NCT03392987)	Treated PS LI MLD (18)	93.8 (95% CI: 63.2 to 99.1) ^h	at 6.0 y of age	-
	NH LI MLD (26)	0 (95% CI: NE)	at 6.0 y of age	
		87.5 (95% CI: 38.7 to 98.1)	at 10.0 y of age	-

	Treated PSEJ MLD (8)	57.1 (95% CI: 21.7 to 81.5) 0 (95% CI: NE)	at 10.0 y of age at 10.0 y of age	- -
	Treated ESEJ MLD (11)			
	NHEJ MLD (17)			
Progression to severe motor impairment or deathⁱ				
Fumagalli 2025 ⁶⁵ (NCT01560182 and NCT03392987)	Treated PSLI MLD (18)	0/18 (0)	at 2 y post-treatment	Absolute difference in proportions, treated PSLI MLD versus NHLI MLD, 60 (95% CI: 36 to 79)
	NHLI MLD (25)	15/25 (60)		
	Treated PSEJ MLD (8)	1/8 (13)	at 2 y post-treatment	Absolute difference in proportions, treated PSEJ MLD versus NHEJ MLD, -13 (95% CI: -53 to 12)
	NHEJ MLD (15)	0/15 (0)		
	Treated ESEJ MLD (11)	2/11 (18)	at 2 y post-treatment	Absolute difference in proportions, treated ESEJ MLD versus NHEJ MLD, -5 (95% CI: -39 to 26)
	NHEJ MLD (15)	2/15 (13)		
	Treated PSLI MLD (13)	2/13 (15)	at 5 y post-treatment	Absolute difference in proportions, treated PSLI MLD versus NHLI MLD, 85 (95% CI: 55 to 98)
	NHLI MLD (26)	26/26 (100)		
	Treated PSEJ MLD (4)	1/4 (25)	at 5 y post-treatment	Absolute difference in proportions, treated PSEJ MLD versus NHEJ MLD, 44 (95% CI: -18 to 82)
	NHEJ MLD (13)	9/13 (69)		
	Treated ESEJ MLD (9)	2/9 (22)	at 5 y post-treatment	Absolute difference in proportions, treated ESEJ MLD versus NHEJ MLD, 69 (95% CI: 25 to 93)
	NHEJ MLD (12)	11/12 (92)		

Groeschel 2016 ³⁰	Treated juvenile MLD (20)	8/20 (40)	at 10 y after onset	RR 0.59 (95% CI: 0.33 to 1.04) ^j Treated juvenile MLD versus control juvenile MLDF, $p = 0.04$
	Control juvenile MLD (41)	28/41 (68)	at 10 y after onset	
Severe cognitive impairment-free survival^k				
Fumagalli 2025 ⁶⁵ (NCT01560182 and NCT03392987)	Treated PS LI MLD (18)	100 (95% CI: 100 to 100) ^l 8.8 (95% CI: 1.1 to 24.3)	at 6.0 y of age at 6.0 y of age	-
	Treated PSEJ MLD (8)	87.5 (95% CI: 38.7 to 98.1) 64.8 (95% CI: 25.3 to 98.1)	at 10.0 y of age at 10.0 y of age	-
	Treated ESEJ MLD (11)	7.5 (95% CI: 0.5 to 28.4)	at 10.0 y of age	-
	NHEJMLD (17)			
No loss of speech				
Fumagalli 2025 ⁶⁵ (NCT01560182 and NCT03392987)	Treated PS LI MLD (17)	100 (95% CI: 100 to 100) ^m 4.2 (95% CI: 0.3 to 17.7)	at 6.0 y of age at 6.0 y of age	-
	Treated PSEJ MLD (7)	100 (95% CI: 100 to 100) 66.7 (95% CI: 28.2 to 87.8)	at 10.0 y of age at 10.0 y of age	-
	Treated ESEJ MLD (11)	26.5 (95% CI: 4.6 to 56.3)	at 10.0 y of age	-
	NHEJMLD (15)			
Language loss				
Groeschel 2016 ³⁰	Treated juvenile MLD (20)	8/20 (40)	at 10 y after onset	RR 0.59 (95% CI: 0.33 to 1.04) ⁱ Treated juvenile MLD versus control juvenile MLDF, $p = 0.07$
	Control juvenile MLD (41)	28/41 (68)	at 10 y after onset	

^a Main natural history cohort (n=26), not including the 2 additional un-treated siblings

^b Four transplant-related mortality; two MLD progression-related mortality.

^c Eleven MLD progression-related mortality.

^d Severe motor impairment defined as GMFC level ≥ 5 (loss of all gross motor function except head control)

^e Severe motor impairment developed in 1 treated patient with PS LI MLD at 7.2 years of age

^f 88% of treated late infantile patients maintained GMFC level 2 (ability to walk) throughout available follow-up.

^g Motor impairment defined as GMFC level ≥ 3 (loss of ability to walk unaided)

^h Motor impairment developed in 1 treated patient with PS LI MLD at 4.4 years of age

ⁱ GMFC level 5 (loss of locomotion and sitting without support).

^j Calculated.

^k Severe cognitive impairment defined as a performance standard score of ≤ 55 and no subsequent scores of > 55

^l Severe motor cognitive developed in 1 treated patient with PS LI MLD at 6.4 years of age

^m One patient with treated patient with PS LI LMD had a loss of speech at 6.4 years of age

CI: confidence interval; EJ: early juvenile; ES: early symptomatic; GMFC: Gross Motor Function Classification; LI: late infantile; MLD: metachromatic leukodystrophy; NE: not evaluable; NH: natural history; NR: not reported; PS: pre-symptomatic RR: relative risk; y: years

Table 11: Continuous outcomes from treatment studies

Publication (Study)	Group (N)	Time point	Mean (SE)	Treatment effect
GMFM total score (%)				
Fumagalli 2025 ⁶⁵ (NCT01560182 and NCT03392987)	Treated PS LI MLD (16) NH LI MLD (11)	2 y after treatment N/A ^a	79.41 (2.64) 8.97 (2.91)	70.44 ^b (95% CI: 62.28 to 78.60),
	Treated PS EJ MLD (7) NH EJ MLD (8)	2 y after treatment N/A ^a	94.27 (0.69) 41.92 (11.34)	52.35 ^b (95% CI: 25.52 to 79.17),
	Treated ES EJ MLD (9) NH EJ MLD (13)	2 y after treatment	86.88 (6.41) 39.64 (9.29)	47.24 ^b (95% CI: 22.86 to 71.62),
	Treated PS LI MLD (8) NH LI MLD (9)	5 y after treatment N/A ^a	79.94 (9.61) 1.26 (0.61)	78.68 ^b (95% CI: 55.96 to 100),
	Treated PS EJ MLD (2) NH EJ MLD (8)	5 y after treatment N/A ^a	99.96 (39.07) 23.99 (8.14)	75.97 ^b (95% CI: -100 to 100),
	Treated ES EJ MLD (3) NH EJ MLD (7)	5 y after treatment N/A ^a	47.34 (13.22) 7.48 (5.06)	39.86 ^b (95% CI: -14.45 to 94.18),
Fumagalli 2022 ²⁹ (NCT01560182)	NH ES EJ MLD (10) NH LI MLD (12)	3 y after treatment N/A ^a	15.9 (NR) 2.8 (NR)	71.5 ^b (95% CI: 50.3 to 92.7), $p = 0.0001$
	Treated EJ MLD (10)	3 y after treatment	72.9 (NR)	

	NHEJMLD (12)	N/A ^a	16.3 (NR)	56.7 ^b (95% CI: 33.7 to 79.6), $p = 0.00061$
	Treated PS EJMLD (4)	3 y after treatment	93.2 (NR)	74.9 ^b (95% CI: 50.8 to 99.1), $p < 0.001$
	NHEJMLD (9)	N/A ^a	18.2 (NR)	
	Treated ES EJMLD (6)	3 y after treatment	59.8 (NR)	43.9 ^b (95% CI: 9.2 to 78.5), $p = 0.054$
	NHEJMLD (10)	N/A ^a	15.9 (NR)	
Brain MRI severity score^c				
Fumagalli 2025 ⁶⁵ (NCT01560182 and NCT03392987)	Treated LIMLD (14)	2 y after treatment	1.63 (0.28)	-9.79 (95% CI: -12.89 to -6.70)
	NHLIMLD (18)	N/A ^a	11.43 (1.43)	
	Treated EJMLD (7)	2 y after treatment	1.83 (0.54)	-10.25 (95% CI: -12.25 to -8.25)
	NHEJMLD (10)	N/A ^a	12.07 (0.75)	
	Treated ES EJMLD (9)	2 y after treatment	8.38 (1.75)	-6.54 (95% CI: -11.72 to -1.35)
	NHEJMLD (13)	N/A ^a	14.91 (1.43)	
	Treated LIMLD (8)	5 y after treatment	2.22 (0.47)	-20.73 (95% CI: -23.71 to -17.75)
	NHLIMLD (8)	N/A ^a	22.94 (1.21)	
	Treated EJMLD (3)	5 y after treatment	2.09 (0.41)	-21.16 (95% CI: -27.05 to -15.28)
	NHEJMLD (6)	N/A ^a	23.25 (2.29)	
	Treated ES EJMLD (3)	5 y after treatment	9.63 (2.86)	-14.63 (95% CI: -26.01 to -3.26)
	NHEJMLD (7)	N/A ^a	24.27 (2.09)	
Fumagalli 2022 ²⁹ (NCT01560182)	Treated LIMLD (8)	3 y after treatment	3.6 (NR)	18.1 ^b (95% CI: 15.0 to 21.1), $p < 0.001$
	NHLIMLD (9)	N/A ^a	21.7 (NR)	
	Treated EJMLD (9)	3 y after treatment	10.1 (NR)	10.4 ^b (95% CI: 3.8 to 21.1), $p = 0.004$
	NHEJMLD (12)	N/A ^a	20.5 (NR)	
Groeschel 2016 ³⁰	Treated juvenile MLD (20)	Last imaging session after HSCT	18.6 (10.2 ^d)	$p = 0.06$
	Control juvenile MLD (36)	NR	22.6 (5.5 ^d)	
NCV index				
Fumagalli 2022 ²⁹	Treated LIMLD (6)	3 y after treatment	-8.3 (NR)	3.2 ^b (95% CI: 1.0 to 5.3),

(NCT01560182) NHLI MLD (8) NA^a -11.5 (NR) $p = 0.01$

^a Age-matched controls.

^b Adjusted MD (covariance model adjusted for treatment and age at assessment).

^c Score ranges from 0 to 31.5, where a score of 0 is considered normal and a score of 31.5 is markedly abnormal

^d SD

CI: confidence interval; EJ: early juvenile; ES: early symptomatic; GMFM: Gross Motor Function Measure; HSCT: hematopoietic stem cell transplantation; LI: late infantile; MD: mean difference; MLD: metachromatic leukodystrophy; MRI: magnetic resonance imaging; N/A: not applicable; NCV: nerve conduction velocity; NH: natural history; NR: not reported; PS pre-symptomatic; SD: standard deviation; SE: standard error; y: years

Table 12: Summary of ROBINS-I assessments

Publication (Study)	Confounding	Selection of participants	Definition of intervention	Risk of Bias			Selection of reported outcome	Overall
				Deviation from intended inter- vention	Missing data	Measurement of outcome		
Fumagalli 2025 ⁶⁵ (NCT01560182 and NCT03392987)	Serious	Low	Low	Low	NI	Serious	Moderate	Serious
Fumagalli 2022 ²⁹ (NCT01560182)	Serious	NI	Low	Low	Serious	Serious	Moderate	Serious
Groeschel 2016 ³⁰	Critical	Serious	Low	NI	Serious	NI	NI	Critical

Low risk of bias - the study is comparable to a well-performed randomised trial with regard to this domain.

Moderate risk of bias - the study is sound for a non-randomised study with regard to this domain but cannot be considered comparable to a well-performed randomised trial.

Serious risk of bias - the study has some important problems in this domain.

Critical risk of bias - the study is too problematic in this domain to provide any useful evidence on the effects of intervention.

No information – no information on which to base a judgement about risk of bias for this domain.

NI: no information

Criterion 14 - Cost effectiveness of NBS for MLD

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

Question 3 - How have modelling studies and cost-effectiveness analyses addressed NBS screening for MLD in the era of novel treatments?

The 2023 UKNSC evidence map⁷ identified five conference abstracts⁶⁸⁻⁷² which were considered relevant to the question:

What is the volume and type of evidence on the cost effectiveness of treatment or screening for MLD in asymptomatic or symptomatic patients?

Four of the five conference abstracts related to cost effectiveness modelling to assess the cost effectiveness of ARSA-cel (Libmeldy®), compared to BSC for the treatment of MLD; studies assessing the cost effectiveness of treatment do not meet the inclusion criteria specified for this evidence summary. The 2023 UKNSC evidence map noted that Bean et al. (2022)⁶⁸ evaluated newborn screening for MLD in the UK and found it to be cost-effective, with an incremental cost-effectiveness ratio (ICER) below £30,000 per quality-adjusted life year (QALY) gained, suggesting a favourable economic case for implementing screening.⁷

The 2023 UKNSC evidence map noted that: ‘the volume and type of evidence on the cost-effectiveness of treatment or screening for MLD is currently limited... further work on the question of cost-effectiveness is also justified.’⁷

What is added by this evidence summary

This evidence summary provides a summary of the published studies available to inform question 3, which includes one full publication, Bean et al. (2024).³¹

Description of new evidence in relation to previous evidence reviews

The searching and title and abstract screening stages of the evidence summary were conducted as a single process, with consideration of all three research questions. Appendix 2 provides an overall PRISMA flow chart for this evidence summary and details of studies included and excluded after full-text screening.

Following full-text screening, there was one publication that met the inclusion criteria specified for research question 3, Bean et al. (2024),³¹ a full publication relating to the study previously reported in the conference abstract Bean et al. (2022).³¹ Bean et al. (2024)³¹ conducted a cost-utility analysis (CUA) using a decision-analytic framework from the perspective of the UK NHS and Personal Social Services (PSS). The model assessed the inclusion of MLD screening in the routine NBS screening programme and was based on a decision-tree model with long-term outcomes estimated using a decision tree for the screening test-related outcomes, and a partitioned survival and state transition model to represent the disease with health states based

on GMFC-MLD. The analysis included 704,328 live births per year in the UK, with epidemiological inputs derived from expert opinion and published literature. Literature sources were used to estimate the screening test positive rates, and clinical expert opinion was used for the decision tree phenotype distribution, no screening arm, and screening test negatives. However, although treatment (ARSA-cel gene therapy [Libmeldy®]) dependent transition probabilities and treatment independent survival curve from the final health state were provided, there was no information on the source of these parameters.

At a 1.5% discount rate, newborn screening for MLD resulted in an incremental gain of 246 QALYs compared with no screening, with an ICER of £33,212 per QALY gained. The study authors concluded that newborn screening for MLD is a cost-effective use of NHS resources using a willingness-to-pay threshold of 50,000/QALY. The main driver of increased costs in the screening arm was the identification and treatment of additional early-onset MLD patients with ARSA-cel gene therapy (Libmeldy®), who would otherwise not be diagnosed in time to receive treatment. Sensitivity analyses explored variations in key parameters, including MLD incidence rates, treatment eligibility probabilities, and discount rates, confirming that newborn screening remained cost-effective under most scenarios. The ICER remained below the £50,000 per QALY willingness-to-pay (WTP) threshold, which the authors concluded supported the inclusion of MLD screening in the UK newborn screening programme as a cost-effective use of NHS resources.

The lead author and three additional study authors were employees of Orchard Therapeutics, the company that manufactures ARSA-cel (Libmeldy®), and two further authors received payment from Orchard Therapeutics for Markov model development.³¹

Methodological quality

The methodological quality of Bean et al. (2024)³¹ was assessed using the Drummond checklist, a tool for evaluating economic evaluations. This checklist is suitable for assessing the study's CUA within a decision-analytic framework. Table 13 provides a summary of the Drummond assessments.

The quality assessment results indicate that Bean et al. (2024)³¹ meets 25 out of 31 applicable criteria. It clearly states the research question, economic importance, and analysis perspective, provides detailed cost and outcome estimates. However, six limitations were identified. The choice of model structure and key parameters lack justification, the discount rate was not well justified, and neither was variable selection or tested ranges in the sensitivity analyses.

Table 13: Summary methodological quality assessment using the Drummond checklist

Drummond checklist	Bean 2024 ³¹
The research question is stated	Yes
The economic importance of the research question is stated	Yes
The viewpoint(s) of the analysis are clearly stated and justified	Yes
The rationale for choosing alternative programmes or interventions compared is stated	Yes
The alternatives being compared are clearly described	Yes
The form of economic evaluation used is stated	Yes
The choice of form of economic evaluation is justified in relation to the questions addressed	Yes

The source(s) of effectiveness estimates used are stated	Yes
Details of the design and results of effectiveness study are given (if based on a single study)	Yes
Details of the methods of synthesis or meta-analysis of estimates are given (if based on a synthesis of a number of effectiveness studies)	Not applicable
The primary outcome measure(s) for the economic evaluation are clearly stated	Yes
Participants who took part in study were representative of whole population?	Yes
Methods to value benefits are stated	Yes
Details of the subjects from whom valuations were obtained were given	Yes
Productivity changes (if included) are reported separately	Not applicable
The relevance of productivity changes to the study question is discussed	Not applicable
Quantities of resource use are reported separately from their unit costs	Yes
Methods for the estimation of quantities and unit costs are described	Yes
Currency and price data are recorded	Yes
Details of currency of price adjustments for inflation or currency conversion are given	Yes
Details of any model used are given	Yes
The choice of model used and the key parameters on which it is based are justified	No
Time horizon of costs and benefits is stated	Yes
The discount rate(s) is stated	Yes
The choice of discount rate(s) is justified	No
An explanation is given if costs and benefits are not discounted	Not applicable
Details of statistical tests and confidence intervals are given for stochastic data	Not clear
The choice of variables for sensitivity analysis is justified	No
The ranges over which the variables are varied are justified	No
Relevant alternatives are compared	Yes
Incremental analysis is reported	Yes
Major outcomes are presented in a disaggregated as well as aggregated form	Yes
The answer to the study question is given	Yes
Conclusions follow from the data reported	Yes
Conclusions are accompanied by the appropriate caveats	No
Yes	25
No	6
Total applicable	3
Not clear	1

Discussion of findings

The available evidence to inform research question 3 ‘How have modelling studies and CEAs addressed newborn screening (NBS) for MLD in the era of novel treatments?’ is derived from a single publication,³¹ which reports an economic evaluation of MLD screening in the UK with substantial methodological limitations.

This evidence summary adds to the existing evidence by including one full publication, Bean et al. (2024),³¹ which provides an economic evaluation of MLD screening in the UK. Bean et al. (2024)³¹ conducted a CUA using a decision-analytic framework from the perspective of the UK NHS and PSS. The analysis included 704,328 live births per year in the UK. The study found that at a 1.5% discount rate, newborn screening for MLD resulted in an incremental gain of 246 QALYs compared with no screening, with an ICER of £33,212 per QALY gained, which was below the authors’ WTP threshold of £50,000 per QALY, suggesting that MLD screening is a cost-effective use of NHS resources.

The findings from Bean et al. (2024)³¹ indicate that newborn screening can significantly increase the number of presymptomatic MLD patients diagnosed within the treatment window, allowing for earlier intervention with ARSA-cel (Libmeldy®), which is associated with substantial improvements in survival and quality of life (QoL).⁵ Sensitivity analyses tested variations in incidence rates, treatment eligibility, and discount rates, demonstrating that NBS remains cost-effective under most scenarios. However, there was a lack of justification in the choice of model parameters and ranges for sensitivity analyses, that could impact result reliability. It is noteworthy that the NICE committee preferred a discount rate of 3.5% to the company's preferred 1.5%, which was the only scenario where the ICER was above £50,000.⁵ Crucially, there was a reliance on clinical expert opinion for many parameters e.g. percentage pre-/early symptomatic early juvenile, and all twelve probabilities for the no screening arm. Clinical expert opinion also informed the probabilities associated with early detection through screening, including the probabilities of progression prior to treatment and on treatment, and the probability of “achieving long term outcomes from treatment”. There was also no information on the source of the treatment effect of ARSA-cel i.e. the transition probabilities with or without treatment, with 100% of patients detected by screening assumed to be full responders to ARSA-cel, as opposed to between 0% and 60% (depending on disease variant) without screening. This means that the robustness of the findings is questionable.

Overall, the findings from Bean et al. (2024)³¹ provide the most comprehensive published economic evaluation of MLD screening to date but potential limitations of the model prevent this evidence summary from drawing reliable conclusions on the case for incorporating MLD into national newborn screening programmes. An External Quality Assurance (EQA) of and changes to the existing model and/or a de novo model would most likely be needed to determine whether UK NSC criterion 14 is met or not met (i.e. criterion 14 is ‘uncertain’). Even with further modelling, if there is no more than clinical expert opinion to inform so many key parameters then further clinical evidence might be required.

Summary of findings relevant to Criterion 14

Following feedback from stakeholders, we have reclassified the findings of this review relating to UK NSC criterion 14, *‘The opportunity cost of the screening programme (including testing, diagnosis and treatment, administration, training and quality assurance) should be economically balanced in relation to expenditure on medical care as a whole (value for money). Assessment against this criterion should have regard to evidence from cost benefit and/or CEAs and have regard to the effective use of available resource’* from ‘not met’ to ‘uncertain’. This change reflects that the format of this evidence summary does not allow the reviewers to determine with certainty whether this criterion is met or not met. Instead, an EQA of and changes to the existing model from Bean et al (2024)³¹, or the creation of a new health economic model, would be needed to provide sufficient evidence to determine if this criterion is met or not met.

It is important to note that the UK NSC is not questioning the cost-effectiveness of Libmeldy when used within its licencing conditions as a treatment for late infantile and early juvenile MLD, as the cost-effectiveness of Libmeldy has already been considered by NICE. This review question is instead examining whether newborn screening followed by treatment is cost-effective compared to treatment without newborn screening.

Review summary

Conclusions and implications for policy

This evidence summary employed standard systematic review methodology to ensure that the capture of relevant evidence was as complete as possible. In addition, to provide further context, information was sought about any existing implemented NBS screening programmes for MLD and any published clinical guidelines on the management of MLD that are relevant to the UK context. We also conducted a horizon scanning exercise to identify any ongoing studies and recent developments in novel therapies for MLD.

Substantial uncertainty remains about the performance of NBS screening programmes for MLD, including with respect to appropriate cut-offs for use in the UK population. There is also uncertainty around the clinical effectiveness of treatments for MLD and, in particular, about whether treatment of pre-symptomatic patients (identified through screening) improves outcomes. Fumagalli et al. (2022)²⁹ and Fumagalli et al. (2025)⁶⁵ provide some evidence that the gene therapy Libmeldy®, recommended in NICE guidance HST18,⁵ may be effective in improving gross motor function for patients with pre-symptomatic/early symptomatic late infantile or early juvenile MLD, relative to an untreated NH cohort.²⁹ There is some very weak, indirect evidence to indicate that the effects of gene therapy treatment (Libmeldy®) on gross motor function, relative to untreated patients, may be greater where patients with early juvenile MLD receive treatment before symptoms develop. This evidence is derived from subgroup analyses within two small studies which reported an indirect comparison with substantial methodological limitations, partially funded by Orchard Therapeutics (the manufacturer of Libmeldy®). The EAG report for NICE guidance HST18⁶⁴ is not included in this evidence summary and most of the clinical effectiveness results included in this report are redacted. However, it should be noted that the key areas of concern raised by the EAG are also applicable to the versions of the analysis published in Fumagalli et al. (2022)²⁹ and Fumagalli et al. (2025)⁶⁵. There is currently no direct evidence that identification of patients with MLD through screening or cascade testing results in improved outcomes.

Criterion 4, *'There should be a simple, safe, precise and validated screening test'* is considered currently not met. This conclusion was based on the lack of data to inform estimates of the accuracy of evaluated screening algorithms and the high uncertainty and variability in estimates of PPV calculated from two studies, for the same 2-tier screening algorithm.¹¹ No studies reporting findings from implemented NBS screening programmes for MLD were identified.

Criterion 5, *'The distribution of test values in the target population should be known and a suitable cut-off level defined and agreed'* was considered currently not met. This conclusion was based on the incidental identification of a new case of late infantile MLD, during the validation phase of a UK pre-pilot study, which raised questions about the appropriate cut-off value for sulfatide levels when used as the 1st tier of a 2-tier screening algorithm.

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'* was considered currently not met. Even with the inclusion of indirect evidence, and the additional consideration of Fumagalli et al (2025)⁶⁵,

evidence about the effectiveness of treatments for MLD is sparse and has substantial methodological limitations. There is currently no direct evidence that identification of patients with MLD through screening or cascade testing results in improved outcomes relative to treatment following symptomatic presentation.

Criterion 14, '*The opportunity cost of the screening programme (including testing, diagnosis and treatment, administration, training and quality assurance) should be economically balanced in relation to expenditure on medical care as a whole (value for money). Assessment against this criterion should have regard to evidence from cost benefit and/or cost effectiveness analyses and have regard to the effective use of available resource*' was considered uncertain. The available evidence was limited to a single publication,³¹ which reports an economic evaluation of MLD screening in the UK; an EQA of and changes to this model and/or a de novo model would most likely be needed to determine whether UK NSC criterion 14 is met or not met.

The current published evidence base alone is not adequate to support implementation of NBS screening for MLD.

We identified only one reference to an implemented screening programme, a news article from Oslo University Hospital, reporting that: '*In January 2025, Norway became the first country in the world to start national screening for metachromatic leukodystrophy (MLD).*'⁴¹ Future publication of findings from the Norwegian programme would further expand the evidence base. One publication¹³ which did not meet the inclusion criteria for this evidence summary, listed ongoing pilot studies and referred to the approval (Illinois, US) of the addition of MLD to the state newborn screening panel, and noted that implementation is anticipated to start in 2024/25. Future publication of data from implemented screening programmes and ongoing pilot studies¹³ has the potential to provide evidence to inform criteria 4 and 5.

Further work is needed to adequately evaluate the performance of screening algorithms for MLD, in practice, and to establish the cut-off values appropriate for use in the UK population. Methodologically robust studies are needed to confirm the clinical effectiveness of available treatments for MLD and to test the hypothesis that treatment outcomes are improved where patients are treated before the onset of symptoms (i.e. through screening). Evidence about the performance of screening algorithms and the efficacy of treatment is a pre-requisite to provide robust model inputs for CEAs.

Limitations

The paucity and poor quality of evidence, across all the criteria considered in this evidence summary, is a key limitation. Evidence generation is still at a relatively early stage and ongoing pilot studies and/or data collection from the first implemented screening programmes are likely to inform future evidence reviews. MLD could be considered as a potential candidate for future evaluation in live NHS services or as part of an international collaboration where the UK is a partner, which may provide further mechanisms to help expand the evidence base.

The systematic review component of this evidence summary was limited by a restriction to full publications in English. This may have resulted in an incomplete picture, particularly in respect of any early screening evaluations/pilot studies conducted internationally.

Appendix 1 - Search strategy

Electronic databases

The searches included the databases shown in Table 14. MEDLINE, MEDLINE In-Process, MEDLINE Daily, Epub Ahead of Print, Embase, CINAHL, Orphanet, Orphanet Newborn Screening, The Cochrane Library (CDSR), CDSR Protocols and CENTRAL, KSR Evidence, ClinicalTrials.gov, EUCTR, WHO ICTRP, ScanMedicine, TRIP Database, GIN, NICE, INAHTA, NIHR HTA, ECRI Guidelines Trust.

Table 14: Databases searched

Resource	Host	Date range	Date searched	Records found
MEDLINE	Ovid	1946 to 2024 October 21	22.10.24	859
Embase	Ovid	1974 to 2024 October 21	22.10.24	1800
CINAHL	EBSCO	2011-2024 October 23	23.10.24	170
Orphanet	https://www.orpha.net	up to 2024 October 22	22.10.24	6
Orphanet Newborn Screening	https://www.orpha.net	up to 2024 October 22	22.10.24	6
CDSR + CDSR Protocols	Wiley	up to 2024 October Iss 10	22.10.24	13
CENTRAL	Wiley	up to 2024 September Iss 9	22.10.24	169
KSR Evidence	https://ksrevidence.com/	up to 2024 October 22	22.10.24	13
Trials Databases				
ClinicalTrials.gov	http://www.clinicaltrials.gov/	up to 2024 October 15	15.10.24	67
EUCTR	https://www.clinicaltrialsregister.eu/	up to 2024 October 21	21.10.24	27
WHO ICTRP	http://www.who.int/ictrp/en/	up to 2024 October 15	15.10.24	45
ScanMedicine	https://scanmedicine.com/	up to 2024 October 21	21.10.24	166
Additional Searches				
TRIP	https://www.tripdatabase.com/	up to 2024 October 21	21.10.24	62
GIN	https://g-i-n.net/international-guidelines-library/	up to 2024 October 21	21.10.24	0
NICE	https://www.nice.org.uk/	up to 2024 October 21	21.10.24	4

International HTA Database	https://database.inahta.org/	up to 2024 October 22	22.10.24	3
NIHR HTA	https://www.nihr.ac.uk/	up to 2024 October 21	21.10.24	7
ECRI	https://home.ecri.org/	up to 2024 October 21	21.10.24	2

CINAHL: Cumulative Index to Nursing and Allied Health Literature; CDSR: Cochrane Database of Systematic Reviews; CENTRAL: Cochrane Central Register of Controlled Trials; ECRI: Europe, Middle East & Africa; EUCTR: EU Clinical Trials Registry; GIN: Guidelines International Network; HTA: Health Technology Assessment; INAHTA: The International HTA Database; KSR: Kleijnen Systematic Reviews Ltd; NICE: National Institute for Health and Care Excellence; NIHR: National Institute for Health and Care Research; WHO ICTRP: World Health Organization International Clinical Trials Registry Platform

Search terms

Search terms included combinations of free text and subject headings (MeSH for MEDLINE, and Emtree terms for Embase), grouped into the following category:

- disease area: MLD

Search terms for MEDLINE, MEDLINE In-Process, MEDLINE Daily, Epub Ahead of Print shown in Table 15, search terms for Embase are shown in Table 16, search terms for CINAHL are shown in Table 17, search terms for Orphanet are shown in Table 18, search terms for Orphanet: Newborn Screening are shown in Table 19, search terms for the Cochrane Library databases are shown in Table 20, search terms for KSR Evidence are shown in Table 21, search terms for ClinicalTrials.gov are shown in Table 22, search terms for EUCTR are shown in Table 23, search terms for WHO ICTRP are shown in Table 24, search terms for ScanMedicine are shown in Table 25, search terms for TRIP are shown in Table 26, search terms for GIN are shown in Table 27, search terms for NICE are shown in Table 28, search terms for the INAHTA are shown in Table 29, search terms for NIHR HTA are shown in Table 30 and search terms for ECRI are shown in Table 31.

Table 15: Search strategy for MEDLINE, MEDLINE In-Process, MEDLINE Daily, Epub Ahead of Print (Searched via Ovid)

Term Group	#	Search terms	Results
Disease	1	Leukodystrophy, Metachromatic/	1347
	2	(MLD and (gene\$ or ARSA or ASA or arylsulfatase or arylsulphatase or leukodystroph\$ or leukodystroph\$)).ti,ab,ot.	1111
	3	(Metachromatic adj2 (leukoencephal\$ or leukoencephal\$ or leukodystroph\$ or leukodystroph\$)).ti,ab,ot,kw,kf,hw.	1874
	4	("Arylsulfatase A" or "arylsulphatase A" or "epididymis secretory sperm binding protein") adj2 deficien\$).ti,ab,ot,kw,kf,hw.	260
	5	Greenfield\$ Disease.ti,ab,ot,kw,kf,hw.	6
	6	(Cerebroside adj2 (Sulfatase or Sulphatase) adj2 Deficien\$).ti,ab,ot,kw,kf,hw.	7
	7	(cerebroside adj2 (sulfate or sulphate) adj2 storage disease).ti,ab,ot,kw,kf,hw.	0
	8	((ASA or ESSPB or ARSA) adj2 Deficien\$).ti,ab,ot,kw,kf,hw.	145
	9	Cerebroside Deficien\$.ti,ab,ot,kw,kf,hw.	0
	10	((diffuse or metachromatic) adj3 (Cerebral or brain) adj3 sclerosis).ti,ab,ot,kw,kf,hw.	2399
	11	((sulfatide or sulphatide) adj2 lipidos).ti,ab,ot,kw,kf,hw.	18
	12	(mckusick-25010 or mckusick25010).ti,ab,ot,kw,kf,hw.	0
	13	(sulfatidosis or sulphatidosis).ti,ab,ot,kw,kf.	18
	14	or/1-13	4662
Limits	15	animals/ not (animals/ and humans/)	5234827

16	14 not 15	4364
17	limit 16 to yr="2012 -Current"	859

Table 16: Search strategy for Embase (Searched via Ovid)

Term Group	#	Search terms	Results
Disease	1	Metachromatic leukodystrophy/	2506
	2	(MLD and (gene\$ or ARSA or ASA or arylsulfatase or arylsulphatase or leukodystroph\$ or leukodystroph\$)).ti,ab,ot.	1863
	3	(Metachromatic adj2 (leukoencephal\$ or leukoencephal\$ or leukodystroph\$ or leukodystroph\$)).ti,ab,ot,kw,hw.	2750
	4	("Arylsulfatase A" or "arylsulphatase A" or "epididymis secretory sperm binding protein") adj2 Deficien\$).ti,ab,ot,kw,hw.	354
	5	Greenfield\$ Disease.ti,ab,ot,kw,hw.	0
	6	(Cerebroside adj2 (Sulfatase or Sulphatase) adj2 Deficien\$).ti,ab,ot,kw,hw.	45
	7	(cerebroside adj2 (sulfate or sulphate) adj2 storage disease).ti,ab,ot,kw,hw.	0
	8	((ASA or ESSPB or ARSA) adj2 Deficien\$).ti,ab,ot,kw,hw.	194
	9	Cerebroside Deficien\$.ti,ab,ot,kw,hw.	0
	10	((diffuse or metachromatic) adj3 (Cerebral or brain) adj3 sclerosis).ti,ab,ot,kw,hw.	29
	11	((sulfatide or sulphatide) adj2 lipidosi\$).ti,ab,ot,kw,hw.	11
	12	(mckusick-25010 or mckusick25010).ti,ab,ot,kw,hw.	0
	13	13 (sulfatidosis or sulphatidosis).ti,ab,ot,kw.	18
	Limits	14	or/1-13
15		animal/	1685351
16		animal experiment/	3223364
17		(rat or rats or mouse or mice or murine or rodent or rodents or hamster or hamsters or pig or pigs or porcine or rabbit or rabbits or animal or animals or dogs or dog or cats or cow or bovine or sheep or ovine or monkey or monkeys).ti,ab,ot,hw.	7972183
18		or/15-17	7972183
19		exp human/	27221576
20		human experiment/	673515
21		or/19-20	27224440
22		18 not (18 and 21)	5924681
23		14 not 22	3490
24		limit 23 to yr="2012 -Current"	1800

Table 17: Search strategy for CINAHL (Searched via EBSCO)

Term Group	#	Search terms	Results
Disease	S1	TI ((MLD and (gene* or ARSA or ASA or arylsulphatase or arylsulphatase or leukodystroph* or leucodystroph*))) OR AB ((MLD and (gene* or ARSA or ASA or arylsulphatase or arylsulphatase or leukodystroph* or leucodystroph*)))	147
	S2	TI ((Metachromatic N2 (leukoencephal* or leukoencephal* or leukodystroph* or leucodystroph*))) OR AB ((Metachromatic N2 (leukoencephal* or leukoencephal* or leukodystroph* or leucodystroph*)))	129
	S3	TI (("Arylsulfatase A" or "arylsulphatase A" or "epididymis secretory sperm binding protein") N2 deficient*) OR AB (("Arylsulfatase A" or "arylsulphatase A" or "epididymis secretory sperm binding protein") N2 deficient*)	10
	S4	TI Greenfield* Disease OR AB Greenfield* Disease	5
	S 5	TI ((Cerebroside N2 (Sulfatase or Sulphatase) N2 Deficien*)) OR AB ((Cerebroside N2 (Sulfatase or Sulphatase) N2 Deficien*))	0
	S 6	TI ((cerebroside N2 (sulfate or sulphate) N2 storage disease)) OR AB ((cerebroside N2 (sulfate or sulphate) N2 storage disease))	0
	S 7	TI (((ASA or ESSPB or ARSA) N2 Deficien*)) OR AB (((ASA or ESSPB or ARSA) N2 Deficien*))	10
	S 8	TI Cerebroside Deficien* OR AB Cerebroside Deficien*	1
	S 9	TI (((diffuse or metachromatic) N3 (Cerebral or brain) N3 sclerosis)) OR AB (((diffuse or metachromatic) N3 (Cerebral or brain) N3 sclerosis))	1
	S 10	TI (((sulfatide or sulphatide) N2 lipidosis)) OR AB (((sulfatide or sulphatide) N2 lipidosis))	0
	S 11	TI ((mckusick-25010 or mckusick25010)) OR AB ((mckusick-25010 or mckusick25010))	0
	S 12	TI ((sulfatidosis or sulphatidosis)) OR AB ((sulfatidosis or sulphatidosis))	0
	S 13	S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12	238
Limits	S 14	S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 Limiters - Publication Date: 20120101-20241231	170

Table 18: Search strategy for Orphanet (Searched via <https://www.orpha.net>)

Term Group	#	Browsed using disease name	Results
Disease	1	MLD	6

Table 19: Search strategy for Orphanet: Newborn Screening Bibliographical Knowledgebase (Searched via <https://nbs.orphanet.app/?lang=en>)

Term Group	#	Search term	Results
Disease	1	Metachromatic leukodystrophy	6

Table 20: Search strategy for CDSR and CENTRAL (Searched via The Cochrane Library [Wiley])

Term Group	#	Search terms	Results
Disease	#1	MeSH descriptor: [Leukodystrophy, Metachromatic] explode all trees	7
	#2	(MLD and (gene* or ARSA or ASA or arylsulfatase or arylsulphatase or leukodystroph* or leukodystroph*)):ti,ab,kw	101
	#3	(Metachromatic near/2 (leukoencephal* or leukoencephal* or leukodystroph* or leukodystroph*)):ti,ab,kw	13
	#4	("Arylsulfatase A" or "Arylsulphatase A" or "epididymis secretory sperm binding protein") near/2 Deficien*):ti,ab,kw	2
	#5	Greenfield* Disease:ti,ab,kw	132
	#6	(Cerebroside near/2 (Sulfatase or Sulphatase) near/2 Deficien*):ti,ab,kw	1
	#7	(cerebroside near/2 (sulfate or sulphate) near/2 storage disease):ti,ab,kw	0
	#8	((ARSA or ASA or ESSBP) near/1 Deficien*):ti,ab,kw	1
	#9	Cerebroside Deficien*):ti,ab,kw	2
	#10	((diffuse or metachromatic) near/3 (Cerebral or brain) near/2 sclerosis):ti,ab,kw	10
	#11	((sulfatide or sulphatide) near/2 lipidosis):ti,ab,kw	0
	#12	(mckusick-25010 or mckusick25010):ti,ab,kw	0
	#13	(sulfatidosis or sulphatidosis):ti,ab,kw	0
	#14	#1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9 or #10 or #11 or #12 or #13 with Cochrane Library publication date Between Jan 2012 and Oct 2024	182

CDSR Retrieved: 11
 CDSRP Retrieved: 2
 CENTRAL Retrieved: 169

Table 21: Search strategy for KSREvidence (searched via <https://ksrevidence.com/>)

Term Group	#	Search terms	Results
Disease	1	(MLD and (gene* or ARSA or ASA or arylsulfatase or arylsulphatase or leukodystroph* or leucodystroph*)) in Title or Abstract	9
	2	(Metachromatic adj2 (leukoencephal* or leukoencephal* or leukodystroph* or leucodystroph*)) in Title or Abstract	6
	3	(("Arylsulfatase A" or "arylsulphatase A" or "epididymis secretory sperm binding protein") adj2 deficient*) in Title or Abstract	2
	4	Greenfield* Disease in Title or Abstract	2
	5	(Cerebroside adj2 (Sulfatase or Sulphatase) adj2 Deficien*) in Title or Abstract	0
	6	(cerebroside adj2 (sulfate or sulphate) adj2 storage disease) in Title or Abstract	0
	7	((ASA or ESSPB or ARSA) adj2 Deficien*) in Title or Abstract	0
	8	Cerebroside Deficien* in Title or Abstract	0
	9	((diffuse or metachromatic) adj3 (Cerebral or brain) adj3 sclerosis) in Title or Abstract	0
	10	((sulfatide or sulphatide) adj2 lipidosis) in Title or Abstract	0
	11	(mckusick-25010 or mckusick25010) in Title or Abstract	0
	12	(sulfatidosis or sulphatidosis) in All text	0
	13	#1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9 or #10 or #11 or #12 in All text	13

Table 22: Search strategy for NIH ClinicalTrials.gov (searched via <http://www.clinicaltrials.gov/>)

Term Group	#	Search terms	Results In condition	Results in Other terms
Disease	1	(sulfatidosis OR sulphatidosis OR mckusick-25010 OR mckusick25010 OR "sulfatide lipidosis" OR "sulphatide lipidosis" OR "diffuse brain sclerosis" OR "metachromatic brain sclerosis" OR "diffuse Cerebral sclerosis" OR "metachromatic Cerebral sclerosis" OR "Cerebroside Deficiency" OR "Cerebroside Deficiencies" OR "Cerebroside Deficient" OR "ARSA Deficiency" OR "ARSA Deficiencies" or	56	67

"ARSA Deficient" OR "ESSBP Deficiency" OR "ESSBP Deficiencies" or "ESSBP Deficient" OR "Cerebroside Sulphatase storage disease" OR "Cerebroside Sulfatase storage disease" OR "Cerebroside Sulfatase Deficiency" OR "Cerebroside Sulphatase Deficiency" OR "Arylsulfatase A Deficiency" OR "Greenfield Disease" OR "Greenfields Disease" OR "Metachromatic leukoencephalopathy" OR "Metachromatic leucoencephalopathy" OR "Metachromatic leukodystrophy" OR "Metachromatic leucodystrophy" OR (MLD AND (gene OR genes OR genetic OR ARSA OR ARSA OR arylsulfatase OR arylsulphatase OR leukodystrophy OR leucodystrophy))	123
Total	123
Total without duplicates	67

Table 23: Search strategy for EU Clinical Trials Register (EUCTR) (Searched via <https://www.clinicaltrialsregister.eu/ctr-search/search>):

Search Interface 1 (pre 2022):

Term Group	#	Search terms	Results
Disease	1	sulfatidosis OR sulphatidosis OR mckusick-25010 OR mckusick25010 OR "sulfatide lipidosis" OR "sulphatide lipidosis" OR "diffuse brain sclerosis" OR "metachromatic brain sclerosis" OR "diffuse Cerebral sclerosis" OR "metachromatic Cerebral sclerosis" OR "Cerebroside Deficiency" OR "Cerebroside Deficiencies" OR "Cerebroside Deficient" OR "ARSA Deficiency" OR "ARSA Deficiencies" or "ARSA Deficient" OR "ESSBP Deficiency" OR "ESSBP Deficiencies" or "ESSBP Deficient" OR "Cerebroside Sulphatase storage disease" OR "Cerebroside Sulfatase storage disease" OR "Cerebroside Sulfatase Deficiency" OR "Cerebroside Sulphatase Deficiency" OR "Arylsulfatase A Deficiency" OR "Greenfield Disease" OR "Greenfields Disease" OR "Metachromatic leukoencephalopathy" OR "Metachromatic leucoencephalopathy" OR "Metachromatic leukodystrophy" OR "Metachromatic	21

leucodystrophy" OR (MLD AND (gene OR genes OR genetic OR ARSA OR ASA OR arylsulfatase OR arylsulphatase OR leukodystrophy OR leucodystrophy))

Search Interface 2: (post 2022)

Term Group	#	Search terms	Results
Disease	1	MLD	6
	2	Metachromatic	0/5
	3	Sulfatidosis	0
	4	Sulphatidosis	0
	5	ESSBP	0
	6	ARSA	0/3
	7	Greenfield	0
	8	Leukodystrophy	0/5
	9	leucodystrophy	0
		Total	19
		Total without duplicates	6

Table 24: Search strategy for World Health Organization International Clinical Trials Registry Platform (WHO ICTRP) (Searched via <http://www.who.int/ictcp/en/>)

Term Group	#	Search terms	Results
Disease	1	sulfatidosis OR sulphatidosis OR mckusick-25010 OR mckusick25010 OR "sulfatide lipidosis" OR "sulphatide lipidosis" OR "diffuse brain sclerosis" OR "metachromatic brain sclerosis" OR "diffuse Cerebral sclerosis" OR "metachromatic Cerebral sclerosis" OR "Cerebroside Deficiency" OR "Cerebroside Deficiencies" OR "Cerebroside Deficient" OR "ARSA Deficiency" OR "ARSA Deficiencies" or "ARSA Deficient" OR "ESSBP Deficiency" OR "ESSBP Deficiencies" or "ESSBP Deficient" OR "Cerebroside Sulphatase storage disease" OR "Cerebroside Sulfatase storage disease" OR "Cerebroside Sulfatase Deficiency" OR "Cerebroside Sulphatase Deficiency" OR "Arylsulfatase A Deficiency" OR "Greenfield Disease" OR "Greenfields Disease" OR "Metachromatic leukoencephalopathy" OR "Metachromatic leucoencephalopathy" OR "Metachromatic leukodystrophy" OR "Metachromatic leucodystrophy	44

2	(MLD AND (gene OR genes OR genetic OR AR SA OR ASA OR arylsulfatase OR arylsulphatase OR leukodystrophy OR leucodystrophy))	26
3	Total	70
4	Total without duplicates	45

Table 25: Search strategy for ScanMedicine (Searched via <https://scanmedicine.com/>)

Term Group	#	Search terms	Results in Trials
Disease	1	sulfatidosis sulphatidosis mckusick-25010 mckusick25010 "sulfatide lipidosis" "sulphatide lipidosis" "diffuse brain sclerosis" "metachromatic brain sclerosis" "diffuse Cerebral sclerosis" "metachromatic Cerebral sclerosis" "Cerebroside Deficiency" "Cerebroside Deficiencies" "Cerebroside Deficient" "ARSA Deficiency" "ARSA Deficiencies" "ARSA Deficient" "ESSBP Deficiency" "ESSBP Deficiencies" "ESSBP Deficient" "Cerebroside Sulphatase storage disease" "Cerebroside Sulfatase storage disease" "Cerebroside Sulfatase Deficiency" "Cerebroside Sulphatase Deficiency" "Arylsulfatase A Deficiency" "Greenfield Disease" "Greenfields Disease" "Metachromatic leukoencephalopathy" "Metachromatic leukoencephalopathy" "Metachromatic leukodystrophy" "Metachromatic leucodystrophy" MLD	166

Table 26: Search strategy for TRIP database (Searched via <https://www.tripdatabase.com/>)

Term Group	#	Search terms	Results In guidelines	Results In Regulatory guidelines
Disease	1	(mld OR metachromatic OR sulfatidosis OR sulphatidosis OR essbp OR arsa OR leukodystrophy OR leucodystrophy OR mckusick25010) from_date:2012 to_date:2024	54	8
		Total		62

Table 27: Search strategy for International Guidelines Library (GIN) (searched via <https://g-i-n.net/international-guidelines-library/>)

Term Group	#	Search terms	Results
Disease	1	(MLD and (gene* or ARSA or ASA or arylsulfatase or arylsulphatase or leukodystroph* or leucodystroph*)) in Title or Abstract	9
	2	(Metachromatic adj2 (leukoencephal* or leucoencephal* or leukodystroph* or leucodystroph*)) in Title or Abstract	6
	3	("Arylsulfatase A" or "arylsulphatase A" or "epididymis secretory sperm binding protein") adj2 deficient*) in Title or Abstract	2
	4	Greenfield* Disease in Title or Abstract	2
	5	(Cerebroside adj2 (Sulfatase or Sulphatase) adj2 Deficien*) in Title or Abstract	0
	6	(cerebroside adj2 (sulfate or sulphate) adj2 storage disease) in Title or Abstract	0
	7	((ASA or ESSPB or ARSA) adj2 Deficien*) in Title or Abstract	0
	8	Cerebroside Deficien* in Title or Abstract	0
	9	((diffuse or metachromatic) adj3 (Cerebral or brain) adj3 sclerosis) in Title or Abstract	0
	10	((sulfatide or sulphatide) adj2 lipidosis) in Title or Abstract	0
	11	(mckusick-25010 or mckusick25010) in Title or Abstract	0
	12	(sulfatidosis or sulphatidosis) in All text	0
	13	#1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9 or #10 or #11 or #12 in All text	13

Table 28: Search strategy for National Institute for Health and Care Excellence (NICE) (Searched via <https://www.nice.org.uk/>)

Term Group	#	Search terms	Results
Disease	1	MLD OR metachromatic OR sulfatidosis OR sulphatidosis OR essbp OR arsa OR leukodystrophy OR leucodystrophy OR mckusick25010	4
		Total	4

Table 29: Search strategy for the International HTA Database (INAHTA) (searched via <https://database.inahta.org/>)

Term Group	#	Search terms	Results
Disease	1	("Leukodystrophy, Metachromatic"[mhe]) OR (sulfatidosis OR sulphatidosis OR mckusick-5010 OR mckusick25010 OR "sulfatide lipidosis" OR "sulphatide lipidosis" OR "diffuse brain sclerosis" OR "metachromatic brain sclerosis" OR "diffuse Cerebral sclerosis" OR "metachromatic Cerebral sclerosis" OR "Cerebroside Deficiency" OR "Cerebroside Deficiencies" OR "Cerebroside Deficient" OR "ARSA A Deficiency" OR "ARSA Deficiencies" or "ARSA Deficient" OR "ESSBP Deficiency" OR "ESSBP Deficiencies" or "ESSBP Deficient" OR "Cerebroside Sulphatase storage disease" OR "Cerebroside Sulfatase storage disease" OR "Cerebroside Sulfatase Deficiency" OR "Cerebroside Sulphatase Deficiency" OR "Arylsulfatase A Deficiency" OR "Greenfield Disease" OR "Greenfields Disease" OR "Metachromatic leukoencephalopathy" OR "Metachromatic leucoencephalopathy" OR "Metachromatic leukodystrophy" OR "Metachromatic leucodystrophy" OR MLD)	3

Table 30: Search strategy for National Institute for Health and Care Research (NIHR) HTA (Searched via <https://www.nihr.ac.uk/>):

Term Group	#	Search terms	Results
Disease	1	mld	0
	2	metachromatic	1
	3	sulfatidosis	0
	4	sulphatidosis	0
	5	essbp	0
	6	arsa	5
	7	leucodystrophy	0
	8	Leukodystrophy	1
	9	mckusick25010	0
		Total	7
		Total after deduplication	5

Table 31: Search strategy for ECRI Guidelines Trust (Searched via <https://guidelines.ecri.org/>)

Term Group	#	Search terms	Results
Disease	1	MLD OR metachromatic OR sulfatidosis OR sulphatidosis OR essbp OR arsa OR leukodystrophy OR leucodystrophy OR mckusick25010	2
		Total	2

Update searches

The searches for the update included searches of the databases shown in Table 32. MEDLINE, MEDLINE In-Process, MEDLINE Daily, Epub Ahead of Print, Embase and medRxiv. Both the Embase and MEDLINE searches were rerun in their entirety (Tables 33 to 35) and deduplicated against the original search results in Endnote.

Table 32: Update Searches

Resource	Host	Date range	Date searched	Records found
MEDLINE	Ovid	1946 to 2025 January 29	30.1.25	879
EMBASE	Ovid	1974 to 2025 January 29	30.1.25	1811
medRxiv	https://www.medrxiv.org/	up to 2025 February 3	3.2.25	77
Total				2767

Table 33: Search strategy for MEDLINE, MEDLINE In-Process, MEDLINE Daily, Epub Ahead of Print (Searched via Ovid)

Term Group	#	Search terms	Results
Disease	1	Leukodystrophy, Metachromatic/	1353
	2	(MLD and (gene\$ or ARSA or ASA or arylsulfatase or arylsulphatase or leukodystroph\$ or leucodystroph\$)).ti,ab,ot	1123
	3	Metachromatic adj2 (leukoencephal\$ or leukoencephal\$ or leukodystroph\$ or leucodystroph\$).ti,ab,ot,kw,kf,hw.	1892
	4	(("Arylsulfatase A" or "arylsulphatase A" or "epididymis secretory sperm binding protein") adj2 deficient\$).ti,ab,ot,kw,kf,hw.	261
	5	Greenfield\$ Disease.ti,ab,ot,kw,kf,hw	6
	6	(Cerebroside adj2 (Sulfatase or Sulphatase) adj2 Deficien\$).ti,ab,ot,kw,kf,hw.	7
	7	(cerebroside adj2 (sulfate or sulphate) adj2 storage disease).ti,ab,ot,kw,kf,hw.	0
	8	((ASA or ESSPB or ARSA) adj2 Deficien\$).ti,ab,ot,kw,kf,hw.	145
	9	Cerebroside Deficien\$.ti,ab,ot,kw,kf,hw.	0

	10	((diffuse or metachromatic) adj3 (Cerebral or brain) adj3 sclerosis).ti,ab,ot,kw,kf,hw.	2399
	11	((sulfatide or sulphatide) adj2 lipidosis).ti,ab,ot,kw,kf,hw.	18
	12	(mckusick-25010 or mckusick25010).ti,ab,ot,kw,kf,hw.	0
	13	(sulfatidosis or sulphatidosis).ti,ab,ot,kw,kf.	18
	14	or/1-13	4683
Limits	15	animals/ not (animals/ and humans/)	5268165
	16	14 not 15	4384
	17	limit 16 to yr="2012 -Current"	879

Table 34: Search strategy for Embase (Searched via Ovid)

Term Group	#	Search terms	Results
Disease	1	Metachromatic leukodystrophy/	2511
	2	(MLD and (gene\$ or ARSA or ASA or arylsulfatase or arylsulphatase or leukodystroph\$ or leukodystroph\$)).ti,ab,ot.	1868
	3	(Metachromatic adj2 (leukoencephal\$ or leukoencephal\$ or leukodystroph\$ or leukodystroph\$)).ti,ab,ot,kw,hw.	2754
	4	("Arylsulfatase A" or "arylsulphatase A" or "epididymis secretory sperm binding protein") adj2 deficient\$.ti,ab,ot,kw,hw.	354
	5	Greenfield\$ Disease.ti,ab,ot,kw,hw.	0
	6	(Cerebroside adj2 (Sulfatase or Sulphatase) adj2 Deficien\$).ti,ab,ot,kw,hw.	44
	7	(cerebroside adj2 (sulfate or sulphate) adj2 storage disease).ti,ab,ot,kw,hw.	0
	8	((ASA or ESSPB or ARSA) adj2 Deficien\$).ti,ab,ot,kw,hw.	196
	9	Cerebroside Deficien\$.ti,ab,ot,kw,hw.	0
	10	((diffuse or metachromatic) adj3 (Cerebral or brain) adj3 sclerosis).ti,ab,ot,kw,hw.	24
	11	((sulfatide or sulphatide) adj2 lipidosis).ti,ab,ot,kw,hw.	11
	12	(mckusick-25010 or mckusick25010).ti,ab,ot,kw,hw.	0
	13	(sulfatidosis or sulphatidosis).ti,ab,ot,kw.	18
	14	or/1-13	3846
Limits	15	animal/	1671863
	16	animal experiment/	3237837
	17	(rat or rats or mouse or mice or murine or rodent or rodents or hamster or hamsters or pig or pigs or porcine or rabbit or rabbits or animal or animals or	7977692

	dogs or dog or cats or cow or bovine or sheep or ovine or monkey or monkeys).ti,ab,ot,hw.	
18	or/15-17	7977692
19	exp human/	27290477
20	human experiment/	672297
21	or/19-20	27293480
22	18 not (18 and 21)	5910654
23	14 not 22	3478
24	limit 23 to yr="2012 -Current"	1811

Table 35: Search strategy for medRxiv (Searched via <https://www.medrxiv.org/>)

Term Group	#	Search terms	Results
Disease	1	Metachromatic leukoencephalopathy	0
	2	Metachromatic leucoencephalopathy	0
	3	(Metachromatic leukodystrophy	0
	4	Metachromatic leucodystrophy	0
	5	MLD	21
	6	Newborn screening	56
		Total	77

Appendix 2 - Included and excluded studies

PRISMA flowchart

Figure 1 summarises the volume of publications included and excluded at each stage of the review. Publications that were included or excluded after the review of full-text articles are detailed below.

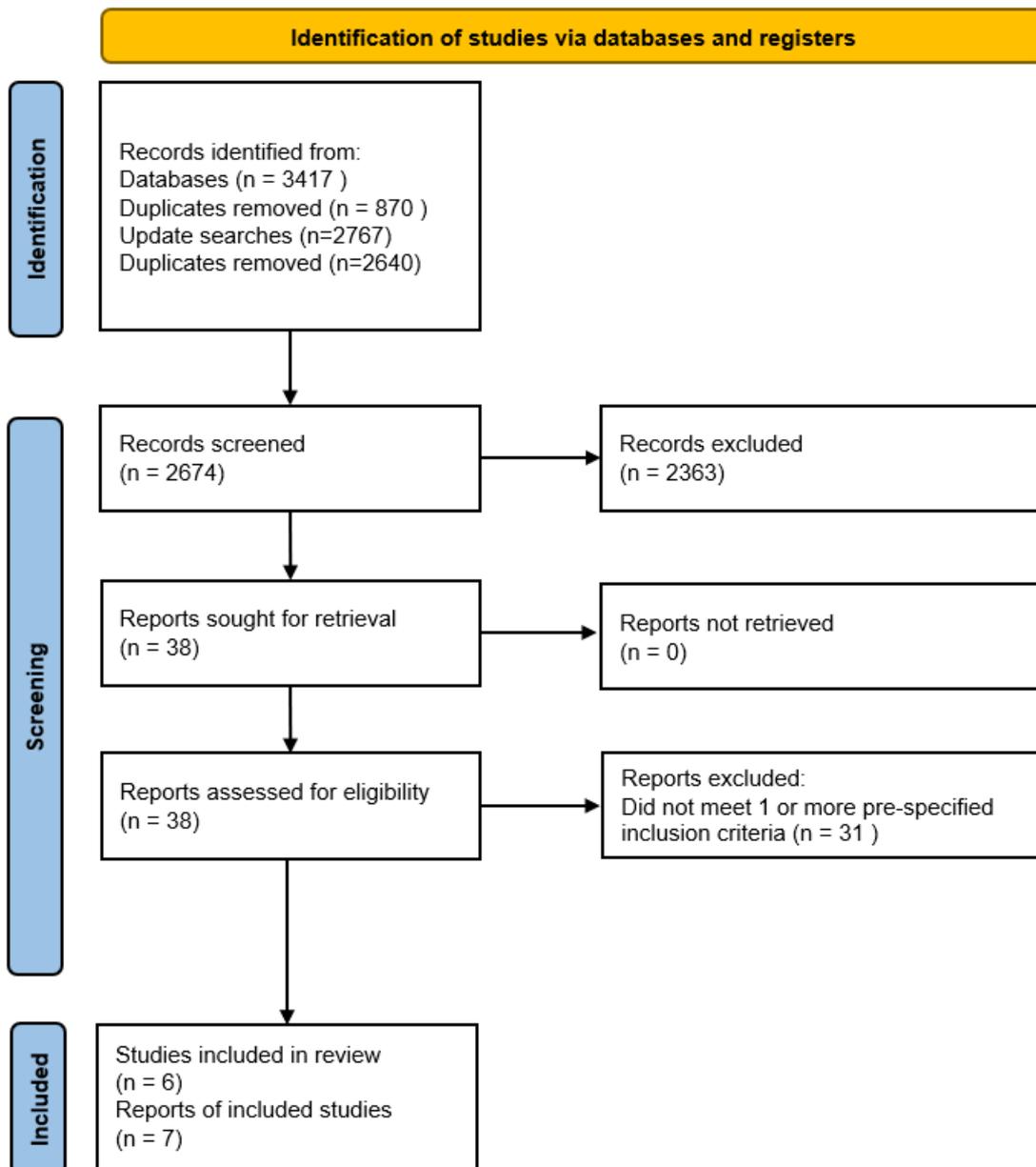


Figure 1: Summary of publications included and excluded at each stage of the review

Publications included after review of full-text articles

The seven publications included after review of full-texts are summarised in Table 36 below.

Table 36: Summary of publications included after review of full-text articles, and the question(s) each publication was identified as being relevant to

Study	The condition	The test	The intervention	The screening programme	Implementation criteria	Comments
Bean, 2024 ³¹				✓		Criterion 14
Fumagalli, 2022 ²⁹			✓			Criterion 9
Groeschel, 2016 ³⁰			✓			Criterion 9
Hong, 2021 ¹¹		✓				Criterion 4
Laugwitz, 2024 ²⁷		✓				Criterion 4
Sessa, 2016 ⁴⁰			✓			Criterion 9
Wu, 2024 ²⁸		✓				Criteria 4 and 5

Publications identified by the 2023 evidence map

The 2023 UKNSC evidence map,⁷ included 25 publications. Details of these publications, whether they were identified by our searches and whether they are included in this evidence summary are provided in Table 37; where studies identified by the 2023 UKNSC evidence map are not included in this evidence summary, reasons for exclusion are provided.

Table 37: Summary of the identification and inclusion or exclusion of publications identified by the 2023 evidence map

Study	Identified by our searches?	Ordered?	Included?	Reason for exclusion
<i>Research question 1 - What is the accuracy of single test and 2-tier NBS screening strategies for MLD, using DBS samples?</i>				
Hong, 2021 ¹¹	Yes	Yes	Yes	-
<i>Research question 2 - Does early initiation of treatment following screening lead to improved outcomes for MLD compared to initiation of treatment following clinical presentation?</i>				
Biffi, 2013 ⁴⁷	Yes	No	No	No comparator
Calbi, 2020 ⁴⁸	Yes	Yes	No	Conference abstract related to an included full paper ²⁹
Calbi, 2023 ⁴⁹	Yes	No	No	Conference abstract and no comparator
Essing, 2021 ⁵⁰	Yes	No	No	Conference abstract and no comparator
Fumagalli, 2017 ⁵¹	Yes	No	No	Conference abstract related to an included full paper ²⁹
Fumagalli, 2019 ⁵²	Yes	Yes	No	Conference abstract and no comparator
Fumagalli, 2019 ⁵³	Yes	No	No	Conference abstract related to an included full paper ²⁹
Fumagalli, 2019 ⁵⁴	Yes	No	No	Conference abstract related to an included full paper ²⁹
Fumagalli, 2020 ⁵⁵	Yes	No	No	Conference abstract related to an included full paper ²⁹

Study	Identified by our searches?	Ordered?	Included?	Reason for exclusion
Fumagalli, 2022 ⁵⁶	Yes	No	No	Conference abstract related to an included full paper ²⁹
Fumagalli, 2022 ²⁹	Yes	Yes	Yes	-
Fumagalli, 2023 ⁵⁷	Yes	No	No	Conference abstract, no full publication identified
Kehrer, 2013 ⁵⁸	Yes	No	No	Conference abstract of a case report
Martin, 2013 ⁵⁹	Yes	No	No	No comparator
Orchard, 2023 ⁶⁰	Yes	No	No	Conference abstract and no comparator
Sessa, 2016 ⁴⁰	Yes	Yes	Yes	-
Sevin, 2018 ⁶¹	Yes	Yes	No	Conference abstract, no full publication identified
Van Rappard, 2016 ⁶²	Yes	Yes	No	Letter to the editor
Yoon, 2020 ⁶³	Yes	Yes	No	Conference abstract, no full publication identified
<i>Research question 3 - How have modelling studies and CEAs addressed NBS screening for MLD in the era of novel treatments</i>				
Bean, 2022 ⁶⁸	Yes	No	No	Excluded (conference abstract); full paper identified and included ³¹
Pang, 2023 ⁷²	Yes	No	No	Conference abstract and cost effectiveness of treatment not screening
Pang, 2022 ⁷⁰	Yes	No	No	Conference abstract and cost effectiveness of treatment not screening
Pang, 2022 ⁶⁹	Yes	No	No	Conference abstract and cost effectiveness of treatment not screening

Study	Identified by our searches?	Ordered?	Included?	Reason for exclusion
Pang, 2021 ⁷¹	Yes	No	No	Conference abstract and cost effectiveness of treatment not screening

Publications excluded after review of full-text articles

Of the 38 publications assessed as potentially relevant after the review of titles and abstracts, 31 were ultimately judged not to be relevant to this review (did not meet the pre-specified inclusion criteria). These publications, along with reasons for exclusion, are listed in Table 38.

Table 38: Publications excluded after review of full-text articles

Publication	Reason for exclusion (PICROS not met)
Barcenas, 2014 ⁷³	Comparison of levels of several types of sulfatide, in DBS and urine samples, from MLD patients versus controls. MLD patients were not newborns and no accuracy related outcome measures were reported or calculable. (P, O, S)
Bekri, 2024 ¹³	Not a primary study; exploration of possible thresholds for various sulfatides as 1st tier tests, reports number (%) above threshold from four pilots (three unpublished). (I, R, O, S)
Bouche, 2015 ⁷⁴	Observational study of long-term outcomes for patients with MLD who have undergone HSCT. States that ' <i>Survival was independent of conditioning regimen, MLD subtype and presence of symptoms at the time of transplantation</i> ', but data were only presented for the comparison of MLD subtypes. (I, C, S)
Calbi, 2020 ⁴⁸	Conference abstract only, no full publication identified.
Calbi, 2023 ⁷⁵	Conference abstract only, no full publication identified.
Calbi, 2024 ⁷⁶	Conference abstract only, no full publication identified.
Chang, 2023 ⁷⁷	Conference abstract only, no full publication identified.
Elmonem, 2014 ⁷⁸	Comparison of the analytical performance of blood spot versus plasma chitotriosidase, using samples from patients with confirmed lysosomal storage disorders and controls; samples were not from newborns and although ROC curves were reported, these were for a combined target condition of 10 different lysosomal storage disorders with no separate data for MLD. (P, O)
Fahim, 2024 ⁷⁹	Not a primary study: summary of a clinical and cost effectiveness assessment from the California Technology Assessment Forum; assessment based on comparison of outcomes for pre-symptomatic and early-symptomatic children with MLD who were treated with atidarsagene versus outcomes in a natural history cohort.
Ferraiuolo, 2012 ⁸⁰	Not a primary study; commentary
Fumagalli, 2019 ⁵²	Conference abstract only, no full publication identified.
Gelb, 2023 ⁸¹	Conference abstract only, no full publication identified.
Hong, 2020 ¹⁰	Assay development paper for the ARSA enzymatic activity assay used in the included study, Hong et al. (2021). Comparison of levels in MLD patients versus healthy adult controls and no accuracy or screening outcomes (levels only). (P, O, S)
Horgan, 2023 ⁸²	Description of RMCH UK patients treated with Libmeldy® since NICE approval: No outcomes and no comparator. (C, O, S)
Jones, 2021 ⁸³	Conference abstract only, no full publication identified.
Jones, 2022 ⁸⁴	Report of an algorithm for assessing screening for IMDs (P, I, C, R, O, S)
Kehrer, 2012 ⁸⁵	Conference abstract only, no full publication identified.
Laugwitz, 2024 ⁴³	Not a primary study; report development of guidelines for clinical management of NBS-identified MLD
Morton, 2022 ⁸⁶	UK and Republic of Ireland survey of caregiver views on early diagnosis and NBS for MLD. (P, I, C, R, O, S)
NICE, 2021 ⁶⁴	Not a primary study: EAG report for NICE HST assessment of Libmeldy®, relevant clinical effectiveness data are redacted.

UKNSC external review — Newborn screening for metachromatic leukodystrophy (MLD), November 2025

Oliva, 2023 ⁸⁷	Conference abstract, reporting data published in included study, Laugwitz et al. (2024)
Pettazzoni, 2023 ⁸⁸	Conference abstract only, no full publication identified.
Ridsdale, 2017 ¹⁷	Conference abstract only, no full publication identified.
Ruis-Schultz, 2021 ⁸⁹	Validation study of an exome sequencing-based NGS method for confirmation testing, using positive NBS samples for a variety of conditions. (P, I, C, R, O, S)
Sevin, 2018 ⁶¹	Conference abstract only, no full publication identified.
Spacil, 2016 ⁸	Assay development paper for the sulfatide assay used in included study, Hong et al. (2021). Comparison of levels in MLD patients versus healthy control newborns, not all MLD patients were newborns and no accuracy outcomes (levels only). (P, O, S)
Suhr, 2017 ⁹⁰	Conference abstract only, no full publication identified.
Van Rappard, 2015 ⁹¹	Conference abstract only, no full publication identified.
Van Rappard, 2016 ⁹¹	Letter to the editor, no full publication identified.
Wiesinger, 2021 ⁹²	Conference abstract only, no full publication identified.
Yoon, 2020 ⁶³	Conference abstract only, no full publication identified.

ARSA: Arylsulfatase A; C: comparator; DBS: dried blood spot; EAG: External Assessment Group; HSCT: hematopoietic stem cell transplantation; HST: Highly Specialised Technology; I: index test or intervention; IMD: inherited metabolic disease; MLD: metachromatic leukodystrophy; NBS: newborn screening; NGS: next generation sequencing; NICE: National Institute for Health and Care Excellence; O: outcomes; P: population; R: reference standard; RMCH: Royal Manchester Children’s Hospital; ROC: receiver operating characteristic; S: study design

Appendix 3 - Appraisal for quality and risk of bias of individual studies

QUADAS-2 and QUADAS C assessments

STUDY: Hong et al. (2021)¹¹ (algorithm a)

DOMAIN 1: PATIENT SELECTION

A. RISK OF BIAS

Stored DBS from de-identified random newborns (n=27,335), provided by the Washington State Department of Health; no further details reported.

Was a consecutive or random sample of patients enrolled? Yes

Was a case-control design avoided? Yes

Did the study avoid inappropriate exclusions? Unclear

Could the selection of patients have introduced bias? RISK: Low

B. APPLICABILITY

No participant details were reported; however, samples were from de-identified random newborns.

Do the included patients match the question? Concerns: Low

DOMAIN 2: INDEX TEST(S)

A. RISK OF BIAS

2-tier screening using two 3 mm punches from the same DBS (as needed). 1st tier testing using UPLC-MS/MS analysis of C16:0-sulfatide levels, with a threshold derived from analysis of stored DBS from 15 known MLD newborns and 2,000 random newborns and set to achieve 100% sensitivity. It was not clear whether the samples from these 2,000 randoms were included in the total of 27,335 samples from random newborns evaluated in the study. 2nd tier testing using ARSA enzyme activity (method and threshold defined in a previous study¹⁰). Confirmatory genetic testing only carried out in screen positive samples and in 3/193 1st tier positive, screen negative samples (process for selecting screen negative samples not reported).

Were the index test results interpreted without knowledge of the results of the reference standard? Yes

If a threshold was used, was it pre-specified? Unclear

Could the conduct or interpretation of the index test have introduced bias? RISK: Unclear

B. APPLICABILITY

Are there concerns that the index test, its conduct, or interpretation differ from the review question? **Concerns: Low**

DOMAIN 3: REFERENCE STANDARD

A. RISK OF BIAS

Confirmatory testing (ARSA exome sequencing) was only undertaken in the screen positive samples and in 3/193 1st tier positive, screen negative samples. It was unclear whether those undertaking ARSA exome sequencing were aware of the results of screening test(s). Given the small numbers of samples involved and the nature of the confirmatory test, it is likely that the results of screening test(s) were known, however, it is not clear whether or how such knowledge could affect interpretation of genetic sequencing.

Is the reference standard likely to correctly classify the target condition? Yes

Were the reference standard results interpreted without knowledge of the results of the index test? Unclear

Could the reference standard, its conduct, or its interpretation have introduced bias? **RISK: Unclear**

B. APPLICABILITY

Is there concern that the target condition as defined by the reference standard does not match the review question? **Concerns: Low**

DOMAIN 4: FLOW AND TIMING

A. RISK OF BIAS

Confirmatory testing (ARSA exome sequencing) was only undertaken in the screen positive samples and in 3/193 1st tier positive, screen negative samples. The process used to select the three selecting screen negative samples for exome sequencing was not reported. No further testing or follow-up was reported for the remaining 190 1st tier positive, screen negative samples, of which 73 did not receive 2nd tier testing due to inadequate sample storage conditions and these samples were excluded from analyses (e.g. reported FPR for 1st tier testing).

Was there an appropriate time interval between index test and reference standard? No

Did patients receive the same or a similar reference standard? No

Were all patients included in the analysis? No

Could the patient flow have introduced bias? **RISK: High**

STUDY: Hong et al. (2021)¹¹ (algorithm b)

DOMAIN 1: PATIENT SELECTION

A. RISK OF BIAS

Stored DBS from de-identified newborns (n=2,287); no further details reported.

Was a consecutive or random sample of patients enrolled? Unclear

Was a case-control design avoided? Yes

Did the study avoid inappropriate exclusions? Unclear

Could the selection of patients have introduced bias? RISK: Unclear

B. APPLICABILITY

No participant details were reported.

Do the included patients match the question? Concerns: Unclear

DOMAIN 2: INDEX TEST(S)

A. RISK OF BIAS

2-tier screening using two 3 mm punches from the same DBS (as needed). 1st tier testing using ARSA enzyme activity (method and threshold defined in a previous study¹⁰). 2nd tier testing using UPLC-MS/MS analysis of C16:0-sulfatide levels, with a threshold derived from analysis of stored DBS from 15 known MLD newborns and 2,000 random newborns and set to achieve 100% sensitivity. It was not clear whether the samples from these 2,000 randoms were included in the total of 2,287 samples from newborns evaluated in the study. No confirmatory genetic testing reported.

Were the index test results interpreted without knowledge of the results of the reference standard? Yes

If a threshold was used, was it pre-specified? Unclear

Could the conduct or interpretation of the index test have introduced bias? RISK: Unclear

B. APPLICABILITY

Are there concerns that the index test, its conduct, or interpretation differ from the review question? Concerns: Low

DOMAIN 3: REFERENCE STANDARD

A. RISK OF BIAS

No confirmatory genetic testing or follow-up was reported.

Is the reference standard likely to correctly classify the target condition? No

Were the reference standard results interpreted without knowledge of the results of the index test? NA

Could the reference standard, its conduct, or its interpretation have introduced bias? RISK: High

B. APPLICABILITY

Is there concern that the target condition as defined by the reference standard does not match the review question? Concerns: High

DOMAIN 4: FLOW AND TIMING

A. RISK OF BIAS

No confirmatory genetic testing or follow-up was reported.

Was there an appropriate time interval between index test and reference standard? No

Did patients receive the same or a similar reference standard? No

Were all patients included in the analysis? Yes

Could the patient flow have introduced bias? RISK: High

STUDY: Hong et al. (2021)¹¹ (QUADAS-C)

DOMAIN 1: PATIENT SELECTION

A. RISK OF BIAS

Was the risk of bias for each index text judged 'low' for this domain?	No
Was a fully paired or randomised design used?	No
Was the sequence allocation random?	NA
Was the allocation sequence concealed until the patients were enrolled and assigned to index tests?	NA

Could the selection of patients have introduced bias in the comparison? RISK: High

DOMAIN 2: INDEX TEST(S)

A. RISK OF BIAS

Was the risk of bias for each index text judged 'low' for this domain?	No
Were index test results interpreted without knowledge of the results of the other index test(s)?	NA
Is undergoing one index test unlikely to affect the performance of the other index tests?	NA
Were the index tests conducted and interpreted without advantaging one of the tests?	Unclear

Could the conduct or interpretation of the index tests have introduced bias in the comparison? RISK: High

DOMAIN 3: REFERENCE STANDARD

A. RISK OF BIAS

Was the risk of bias for each index text judged 'low' for this domain?	No
Did the reference standard avoid incorporating any of the index tests?	Yes

Could the reference standard, its conduct, or its interpretation have introduced bias? RISK: High

DOMAIN 4: FLOW AND TIMING

A. RISK OF BIAS

Was the risk of bias for each index text judged 'low' for this domain? No

Was there an appropriate interval between the tests? NA

Was the same reference standard used for all index tests? No

Are the proportions and reasons for missing data similar across index tests? No

Could the patient flow have introduced bias in the comparison? RISK: High

Laugwitz et al. (2024)²⁷

DOMAIN 1: PATIENT SELECTION

A. RISK OF BIAS

A newborn screening pilot programme, conducted as a prospective cohort, which included newborns in hospitals referring to the newborn screening laboratory in Hannover, Germany. Newborns were included where there was consent from the legal guardian, where a DBS sample had been collected within the first 36 to 72 hours of life (in line with German national guidelines) and where there was sufficient residual DBS sample after completion of the regular national NBS programme.

Was a consecutive or random sample of patients enrolled? Yes

Was a case-control design avoided? Yes

Did the study avoid inappropriate exclusions? Yes

Could the selection of patients have introduced bias? RISK: Low

B. APPLICABILITY

Newborns in an unselected population screening pilot study.

Do the included patients match the question? Concerns: Low

DOMAIN 2: INDEX TEST(S)

A. RISK OF BIAS

3-tier screening using 3.2 mm punch from a DBS. 1st tier testing using UPLC-MS/MS analysis to quantify the sulfatide species C16:0, C16:0-OH and C16:1-OH, preliminary cut-offs were established during validation using 500 random DBS samples and five DBS from symptomatic children with MLD, cut-offs were adjusted over time with the final cut-offs established after screening 109,259 newborns (cut-offs were ≥ 0.17 and ≥ 0.050 $\mu\text{mol/L}$ or ≥ 1.83 and ≥ 3.13 as MoM for C16:0 and C16:1-OH, respectively). 2nd tier testing using ARSA enzyme activity using LC-MS/MS (published method¹⁰), cut-off ≤ 0.015 $\mu\text{mol/L/h}$. 3rd tier genetic testing in DBS clinically relevant variants in ARSA if detected as one homozygous or a combination of two heterozygous variants, biallelic variants in SUMF1 or PSAP (positive screening results for MSD or Prosaposin B deficiency) were not considered FP, but were excluded from reporting, 3rd tier testing done in all DBS with elevated sulfatides (i.e. all 1st tier positives). Concerns about applicability arise from inclusion of genetic testing (more usually considered part of confirmatory testing) in the screening algorithm.

Were the index test results interpreted without knowledge of the results of the reference standard? Yes

If a threshold was used, was it pre-specified? No

Could the conduct or interpretation of the index test have introduced bias? RISK: High

B. APPLICABILITY

Are there concerns that the index test, its conduct, or interpretation differ from the review question? **Concerns: High**

DOMAIN 3: REFERENCE STANDARD

A. RISK OF BIAS

Screen positives (following 3rd tier testing) were referred for confirmatory diagnostics, including urinary sulfatides, ARSA enzyme activity in leukocytes and genetic sequencing of index case and parents (NGS or Sanger sequencing, with variants classified by two geneticists independently based on ACMG guidelines). No testing or follow-up of screen negative patients was reported and no population surveillance to identify any possible screen negative cases was reported. Low risk of bias rating because knowledge of screening results is unlikely to bias genetic sequencing and interpretation by to independent geneticists.

Is the reference standard likely to correctly classify the target condition? **Yes**

Were the reference standard results interpreted without knowledge of the results of the index test? **Unclear**

Could the reference standard, its conduct, or its interpretation have introduced bias? **RISK: Low**

B. APPLICABILITY

Is there concern that the target condition as defined by the reference standard does not match the review question? **Concerns: Low**

DOMAIN 4: FLOW AND TIMING

A. RISK OF BIAS

230/381 1st tier positive samples received 2nd tier screening test and all 381 1st tier positive samples received 3rd tier screening test. Only 3rd tier screen positive patients received confirmatory diagnostic testing. No testing or follow-up of screen negative patients was reported and no population surveillance to identify any possible screen negative cases was reported.

Was there an appropriate time interval between index test and reference standard? **No**

Did patients receive the same or a similar reference standard? **No**

Were all patients included in the analysis? **No**

Could the patient flow have introduced bias? **RISK: High**

STUDY: Wu et al. (2024)²⁸

DOMAIN 1: PATIENT SELECTION

A. RISK OF BIAS

De-identified DBS from the UK newborn screening programme (Manchester Newborn Screening Laboratory). Blood spots collected from babies ≤ 4 days or >12 months of age, rejected due to blood transfusion, or of poor quality; parent declined any research being performed on the baby's residual sample. No further details were reported.

Was a consecutive or random sample of patients enrolled? Unclear

Was a case-control design avoided? Yes

Did the study avoid inappropriate exclusions? Unclear

Could the selection of patients have introduced bias? RISK: Unclear

B. APPLICABILITY

No participant details were reported.

Do the included patients match the question? Concerns: Unclear

DOMAIN 2: INDEX TEST(S)

A. RISK OF BIAS

2-tier screening using DBS. 1st tier testing using UPLC-MS/MS analysis of C16:0-sulfatide levels, using a published threshold¹¹ (≥ 0.17 $\mu\text{mol/L}$), which had been assessed in a validation phase before the start of the pre-pilot (using different samples). 2nd tier testing using ARSA enzyme activity $<20\%$ of mean (method and threshold defined in previous studies^{10, 11}). Confirmatory genetic testing only carried out in 1st tier screen positive samples (there were no 2nd tier screen positives) and in two additional samples with ARSA activity $<20\%$ of mean, identified in the validation phase.

Were the index test results interpreted without knowledge of the results of the reference standard? Yes

If a threshold was used, was it pre-specified? Yes

Could the conduct or interpretation of the index test have introduced bias? RISK: Low

B. APPLICABILITY

Are there concerns that the index test, its conduct, or interpretation differ from the review question? Concerns: Low

DOMAIN 3: REFERENCE STANDARD

A. RISK OF BIAS

PCR amplification of ARSA gene from DBS using an automated EZ1 DNA Tissue Kit (QIAgen) followed by Sanger sequencing; conducted in the 11 1st tier positive, 2nd tier negative samples from the pre-pilot study and in two samples from the initial validation of the ARSA activity reference ranges with ARSA activity <20% of the mean. No testing or follow-up of screen negative patients was reported and no population surveillance to identify any possible screen negative cases was reported. Given the small numbers of samples involved and the nature of the confirmatory test, it is likely that the results of screening test(s) were known, however, it is not clear whether or how such knowledge could affect interpretation of genetic sequencing.

Is the reference standard likely to correctly classify the target condition? Yes

Were the reference standard results interpreted without knowledge of the results of the index test? Unclear

Could the reference standard, its conduct, or its interpretation have introduced bias? RISK: Unclear

B. APPLICABILITY

Is there concern that the target condition as defined by the reference standard does not match the review question? Concerns: Low

DOMAIN 4: FLOW AND TIMING

A. RISK OF BIAS

Confirmatory testing (genetic sequencing to identify pathogenic ARSA variants) was only conducted in the 11 1st tier positive, 2nd tier negative samples from the pre-pilot study and in two samples from the initial validation of the ARSA activity reference ranges with ARSA activity <20% of the mean. No testing or follow-up of screen negative patients was reported and no population surveillance to identify any possible screen negative cases was reported.

Was there an appropriate time interval between index test and reference standard? No

Did patients receive the same or a similar reference standard? No

Were all patients included in the analysis? Yes

Could the patient flow have introduced bias? RISK: High

ROBINS-I assessments

Responses underlined in green are potential markers for low risk of bias, and responses in **red** are potential markers for a risk of bias. Where questions relate only to sign posts to other questions, no formatting is used.

Risk of bias ratings are defined as follows:³⁷

Low risk of bias - the study is comparable to a well-performed randomised trial with regard to this domain

Moderate risk of bias - the study is sound for a non-randomised study with regard to this domain but cannot be considered comparable to a well-performed randomised trial

Serious risk of bias - the study has some important problems in this domain

Critical risk of bias - the study is too problematic in this domain to provide any useful evidence on the effects of intervention

No information – no information on which to base a judgement about risk of bias for this domain

STUDY: Fumagalli et al. (2025)⁶⁵

Signalling questions	Description	Response options
Bias due to confounding		
1.1 Is there potential for confounding of the effect of intervention in this study? If <u>N/PN</u> to 1.1: the study can be considered to be at low risk of bias due to confounding and no further signalling questions need be considered If Y/PY to 1.1: determine whether there is a need to assess time-varying confounding:	Treated patients were compared with patients from a natural history cohort. Matching considered only participant age and disease subtype (no matching by symptom status or severity at baseline).	Y
1.2. Was the analysis based on splitting participants' follow up time according to intervention received? If <u>N/PN</u>, answer questions relating to baseline confounding (1.4 to 1.6) If Y/PY, go to question 1.3.	Analysis compared treated patients, from a clinical trial and expanded access schemes, to a separate natural history cohort of untreated patients.	N

<p>1.3. Were intervention discontinuations or switches likely to be related to factors that are prognostic for the outcome? If N/PN, answer questions relating to baseline confounding (1.4 to 1.6) If Y/PY, answer questions relating to both baseline and time-varying confounding (1.7 and 1.8)</p>	<p>Lentiviral haematopoietic stem cell gene therapy is a single intervention (discontinuation/switching not applicable).</p>	<p>N</p>
--	--	----------

<p>Questions relating to baseline confounding only</p>		
<p>1.4. Did the authors use an appropriate analysis method that controlled for all the important confounding domains?</p>	<p>Gross Motor Function Measure scores for treated patients were compared with natural history patients matched by age and disease subtype using an analysis of covariance model adjusted for treatment and age at assessment of gross motor function.</p>	<p>N</p>
<p>1.5. If Y/PY to 1.4: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?</p>		<p>NA</p>
<p>1.6. Did the authors control for any post-intervention variables that could have been affected by the intervention?</p>		<p>PN</p>
<p>Questions relating to baseline and time-varying confounding</p>		
<p>1.7. Did the authors use an appropriate analysis method that controlled for all the important confounding domains and for time-varying confounding?</p>		<p>NA</p>
<p>1.8. If Y/PY to 1.7: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?</p>		<p>NA</p>
<p>Risk of bias judgement</p>		<p>Serious</p>

Bias in selection of participants into the study

<p>2.1. Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention? If <u>N/PN</u> to 2.1: go to 2.4</p> <p>2.2. If <u>Y/PY</u> to 2.1: Were the post-intervention variables that influenced selection likely to be associated with intervention?</p> <p>2.3 If <u>Y/PY</u> to 2.2: Were the post-intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome?</p>	<p>The treated cohort comprised 39 patients (20 patients from NC T01560182, 10 patients from NCT03392987 and 9 patients treated under expanded access schemes). Two patients (both from NCT01560182) were excluded from the integrated efficacy analysis following adjudication by the IRC as symptomatic LI MLD (1 patient) and progressively symptomatic EJ MLD (1 patient); these patients had evidence of rapid disease progression during the period shortly before treatment and would not therefore meet current eligibility criteria for treatment.</p> <p>The natural history cohort included 43 MLD patients from a natural history study,⁹³ of whom 21 were untreated siblings of patients in the treated cohort, and 6 additional siblings of treated patients.</p>	<p><u>PN</u></p> <p>NA</p> <p>NA</p>
<p>2.4. Do start of follow-up and start of intervention coincide for most participants?</p>		<p><u>PY</u></p>
<p>2.5. If <u>Y/PY</u> to 2.2 and 2.3, or <u>N/PN</u> to 2.4: Were adjustment techniques used that are likely to correct for the presence of selection biases?</p>		<p>NA</p>
<p>Risk of bias judgement</p>		<p>Low</p>

<p>Bias in classification of interventions</p>		
<p>3.1 Were intervention groups clearly defined?</p>	<p>Intervention group from a single arm clinical trial and expanded access schemes, 'control' group was a separate, untreated natural history cohort.</p>	<p><u>Y</u></p>
<p>3.2 Was the information used to define intervention groups recorded at the start of the intervention?</p>		<p><u>Y</u></p>
<p>3.3 Could classification of intervention status have been affected by knowledge of the outcome or risk of the outcome?</p>		<p><u>N</u></p>
<p>Risk of bias judgement</p>		<p>Low</p>

Bias due to deviations from intended interventions		
If your aim for this study is to assess the effect of assignment to intervention, answer questions 4.1 and 4.2		
4.1. Were there deviations from the intended intervention beyond what would be expected in usual practice?	Protocol variations relating to treatment (haematopoietic stem cell transplant procedure) were reported for three patients. ²⁹	PN
4.2. If Y/PY to 4.1: Were these deviations from intended intervention unbalanced between groups <i>and</i> likely to have affected the outcome?		NA
If your aim for this study is to assess the effect of starting and adhering to intervention, answer questions 4.3 to 4.6		
4.3. Were important co-interventions balanced across intervention groups?		NA
4.4. Was the intervention implemented successfully for most participants?		NA
4.5. Did study participants adhere to the assigned intervention regimen?		NA
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
Risk of bias judgement		Low

Bias due to missing data		
5.1 Were outcome data available for all, or nearly all, participants?	Two patients were excluded from the integrated efficacy analysis following adjudication by the IRC as symptomatic LI MLD (1 patient and progressively symptomatic EJ MLD (1 patient); these patients had evidence of rapid disease progression during the period shortly before treatment and would not therefore meet current eligibility criteria for treatment. These patients were described as having experienced motor and cognitive decline at similar rates to those of natural history patients.	Y
5.2 Were participants excluded due to missing data on intervention status?		N
5.3 Were participants excluded due to missing data on other variables needed for the analysis?		PY
5.4 If PNN 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

<p>5.5 If PNN 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?</p>	<p>If a patient withdrew from the study prior to experiencing an event, their time to event was censored at the date of last assessment or the date of last contact or death. For categoric endpoints, the incidence of the event was based on the observed components, and patients who withdrew from the study prior to the analysis time point and had not experienced the event were not included in the analysis. Sensitivity analyses were conducted imputing all missing data following withdrawal as events and imputing all missing data following withdrawal as non-events. For continuous endpoints, analyses are based on data for patients who had reached the analysis time point, and patients who had previously withdrawn were not included in the analysis. Supplementary analyses were conducted after imputing a worst-case value for these patients. The article states that conclusions from all sensitivity and supplementary analyses were consistent with those of primary analyses (results not reported).</p>	<p>NI</p>
<p>Risk of bias judgement</p>		<p>NI</p>

<p>Bias in measurement of outcomes</p>		
<p>6.1 Could the outcome measure have been influenced by knowledge of the intervention received?</p>	<p>The clinical trial was open label and some measures of function used involve subjective assessment.</p>	<p>PY</p>
<p>6.2 Were outcome assessors aware of the intervention received by study participants?</p>		<p>Y</p>
<p>6.3 Were the methods of outcome assessment comparable across intervention groups?</p>	<p>The same measurement scales were used, but no details were reported regarding how these were applied (particularly with respect to patients in the natural history cohort).</p>	<p>NI</p>
<p>6.4 Were any systematic errors in measurement of the outcome related to intervention received?</p>		<p>NI</p>
<p>Risk of bias judgement</p>		<p>Serious</p>

<p>Bias in selection of the reported result</p>		
<p>Is the reported effect estimate likely to be selected, on the basis of the results, from...</p>		

UKNSC external review — Newborn screening for metachromatic leukodystrophy (MLD), November 2025

<p>7.1. ... multiple outcome <i>measurements</i> within the outcome domain?</p>	<p>The article reports the primary outcome measure as survival free from severe motor impairment, defined as the time from birth to the first occurrence of loss of locomotion and of sitting without support (GMFC-MLD level of ≥ 5) or death from any cause. This differs from the primary outcome measures specified for each of the two trial from which treated patents were taken.</p> <p>NCT01560182 - improvement of >10% in the total Gross Motor Function Measure score at 2-years.</p> <p>NCT03392987 - GMFM to evaluate subject's ability to perform specific tasks in different position at 24-months post-treatment.</p>	<p>PY</p>
<p>7.2 ... multiple <i>analyses</i> of the intervention-outcome relationship?</p>		<p>PN</p>
<p>7.3 ... different <i>subgroups</i>?</p>		<p>PN</p>
<p>Risk of bias judgement</p>		<p>Moderate</p>

<p>Overall bias</p>		
<p>Risk of bias judgement</p>	<p>There are substantial issues around inadequate consideration of potential confounders, missing data and possible selective reporting of outcomes.</p>	<p>Serious</p>
<p>N: no; NA: not applicable; NI: no information; PN: probably no; PY: probably yes; Y: yes</p>		

STUDY: Fumagalli et al. (2022)²⁹

Signalling questions	Description	Response options
Bias due to confounding		
<p>1.1 Is there potential for confounding of the effect of intervention in this study? If N/PN to 1.1: the study can be considered to be at low risk of bias due to confounding and no further signalling questions need be considered If Y/PY to 1.1: determine whether there is a need to assess time-varying confounding:</p>	<p>Treated patients were compared with patients from a natural history cohort. Matching considered only participant age and disease subtype (no matching by symptom status or severity at baseline).</p>	<p style="text-align: center;">Y</p>
<p>1.2. Was the analysis based on splitting participants' follow up time according to intervention received? If N/PN, answer questions relating to baseline confounding (1.4 to 1.6) If Y/PY, go to question 1.3.</p>	<p>Analysis compared treated patients, from a clinical trial and expanded access schemes, to a separate natural history cohort of untreated patients.</p>	<p style="text-align: center;">N</p>
<p>1.3. Were intervention discontinuations or switches likely to be related to factors that are prognostic for the outcome? If N/PN, answer questions relating to baseline confounding (1.4 to 1.6) If Y/PY, answer questions relating to both baseline and time-varying confounding (1.7 and 1.8)</p>	<p>Lentiviral haematopoietic stem cell gene therapy is a single intervention (discontinuation/switching not applicable).</p>	<p style="text-align: center;">N</p>
Questions relating to baseline confounding only		
<p>1.4. Did the authors use an appropriate analysis method that controlled for all the important confounding domains?</p>	<p>Gross Motor Function Measure scores for treated patients were compared with natural history patients matched by age and disease subtype using an analysis of covariance model adjusted for treatment and age at assessment of gross motor function.</p>	<p style="text-align: center;">N</p>
<p>1.5. If Y/PY to 1.4: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?</p>		<p style="text-align: center;">NA</p>

1.6. Did the authors control for any post-intervention variables that could have been affected by the intervention?		PN
Questions relating to baseline and time-varying confounding		
1.7. Did the authors use an appropriate analysis method that controlled for all the important confounding domains and for time-varying confounding?		NA
1.8. If Y/PY to 1.7: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?		NA
Risk of bias judgement		Serious

Bias in selection of participants into the study		
2.1. Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention? If N/PN to 2.1: go to 2.4	No details provided regarding the selection of the natural history patients.	NI
2.2. If Y/PY to 2.1: Were the post-intervention variables that influenced selection likely to be associated with intervention?		NI
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?		NI
2.4. Do start of follow-up and start of intervention coincide for most participants?		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?		NA
Risk of bias judgement		NI

Bias in classification of interventions		
3.1 Were intervention groups clearly defined?		Y

3.2 Was the information used to define intervention groups recorded at the start of the intervention?	Intervention group from a single arm clinical trial and expanded access schemes, 'control' group was a separate, untreated natural history cohort.	<u>Y</u>
3.3 Could classification of intervention status have been affected by knowledge of the outcome or risk of the outcome?		<u>N</u>
Risk of bias judgement		Low

Bias due to deviations from intended interventions		
If your aim for this study is to assess the effect of assignment to intervention, answer questions 4.1 and 4.2		
4.1. Were there deviations from the intended intervention beyond what would be expected in usual practice?	Protocol variations relating to treatment (haematopoietic stem cell transplant procedure) were reported for three patients.	<u>PN</u>
4.2. If Y/PY to 4.1: Were these deviations from intended intervention unbalanced between groups <i>and</i> likely to have affected the outcome?		NA
If your aim for this study is to assess the effect of starting and adhering to intervention, answer questions 4.3 to 4.6		
4.3. Were important co-interventions balanced across intervention groups?		NA
4.4. Was the intervention implemented successfully for most participants?		NA
4.5. Did study participants adhere to the assigned intervention regimen?		NA
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
Risk of bias judgement		Low

Bias due to missing data

UKNSC external review — Newborn screening for metachromatic leukodystrophy (MLD), November 2025

5.1 Were outcome data available for all, or nearly all, participants?	Five patients with late infantile MLD did not contribute data to the co-primary analysis (Gross Motor Function Measure at 2-years) because they missed (n=1) or did not reach (n=4) their 2-year assessment. Eight patients with late infantile MLD and one patient with early juvenile MLD in the natural history cohort were not included in the matched analysis because their gross motor function assessments were performed at an age above the upper limit for this analysis.	N
5.2 Were participants excluded due to missing data on intervention status?		N
5.3 Were participants excluded due to missing data on other variables needed for the analysis?		Y
5.4 If PNN 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?		PN
5.5 If PNN 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?		N
Risk of bias judgement		Serious

Bias in measurement of outcomes		
6.1 Could the outcome measure have been influenced by knowledge of the intervention received?	The clinical trial was open label and some measures of function used involve subjective assessment.	PY
6.2 Were outcome assessors aware of the intervention received by study participants?		Y
6.3 Were the methods of outcome assessment comparable across intervention groups?	The same measurement scales were used, but no details were reported regarding how these were applied (particularly with respect to patients in the natural history cohort).	NI
6.4 Were any systematic errors in measurement of the outcome related to intervention received?		NI
Risk of bias judgement		Serious

Bias in selection of the reported result		
Is the reported effect estimate likely to be selected, on the basis of the results, from...		

UKNSC external review — Newborn screening for metachromatic leukodystrophy (MLD), November 2025

7.1. ... multiple outcome <i>measurements</i> within the outcome domain?	A (clinically relevant) improvement of >10% in the total Gross Motor Function Measure score at 2-years was cited as a co-primary endpoint but was not reported; outcomes reported were total Gross Motor Function Measure score and severe motor impairment-free survival.	PY
7.2 ... multiple <i>analyses</i> of the intervention-outcome relationship?		PN
7.3 ... different <i>subgroups</i> ?		PN
Risk of bias judgement		Moderate

Overall bias		
Risk of bias judgement	There are substantial issues around inadequate consideration of potential confounders, missing data and possible selective reporting of outcomes.	Serious
N: no; NA: not applicable; NI: no information; PN: probably no; PY: probably yes; Y: yes		

STUDY: Groeschel et al. (2016)³⁰

Signalling questions	Description	Response options
Bias due to confounding		
<p>1.1 Is there potential for confounding of the effect of intervention in this study? If N/PN to 1.1: the study can be considered to be at low risk of bias due to confounding and no further signalling questions need be considered</p>	<p>Long-term outcomes were compared between patients with juvenile MLD born between 1975 and 2009 and who received haematopoietic stem cell therapy at a median age of 7 years (age range, 1.5-18.2 years) and non-transplanted patients with juvenile MLD born between 1967 and 2007. No matching of treated and untreated patients or consideration of potential confounders was reported.</p>	<p style="text-align: center;">Y</p>
<p>If Y/PY to 1.1: determine whether there is a need to assess time-varying confounding:</p>		
<p>1.2. Was the analysis based on splitting participants' follow up time according to intervention received? If N/PN, answer questions relating to baseline confounding (1.4 to 1.6) If Y/PY, go to question 1.3.</p>		<p style="text-align: center;">N</p>
<p>1.3. Were intervention discontinuations or switches likely to be related to factors that are prognostic for the outcome? If N/PN, answer questions relating to baseline confounding (1.4 to 1.6) If Y/PY, answer questions relating to both baseline and time-varying confounding (1.7 and 1.8)</p>		<p style="text-align: center;">NA</p>
Questions relating to baseline confounding only		
<p>1.4. Did the authors use an appropriate analysis method that controlled for all the important confounding domains?</p>	<p>No consideration of potential confounders was reported.</p>	<p style="text-align: center;">N</p>
<p>1.5. If Y/PY to 1.4: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?</p>		<p style="text-align: center;">NA</p>

1.6. Did the authors control for any post-intervention variables that could have been affected by the intervention?		PN
Questions relating to baseline and time-varying confounding		
1.7. Did the authors use an appropriate analysis method that controlled for all the important confounding domains and for time-varying confounding?	Simple comparisons between treated and 'control' patients; no consideration of confounding.	N
1.8. If Y/PY to 1.7: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?		NA
Risk of bias judgement		Critical

Bias in selection of participants into the study		
2.1. Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention? If N/PN to 2.1: go to 2.4	Patients followed up for at least 2 years after transplantation were included from three German centres performing haematopoietic stem cell therapy in patients with MLD. Non-transplanted control patients were recruited between 2006 and 2014 as part of a NH study within the German leukodystrophy network Leukonet.	Y
2.2. If Y/PY to 2.1: Were the post-intervention variables that influenced selection likely to be associated with intervention?		PY
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?		PY
2.4. Do start of follow-up and start of intervention coincide for most participants?		<u>PY</u>
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
Risk of bias judgement		Serious

Bias in classification of interventions		
3.1 Were intervention groups clearly defined?	Patients who had received haematopoietic stem cell therapy were compared to non-transplanted controls.	<u>Y</u>

3.2 Was the information used to define intervention groups recorded at the start of the intervention?		<u>Y</u>
3.3 Could classification of intervention status have been affected by knowledge of the outcome or risk of the outcome?		<u>N</u>
Risk of bias judgement		Low

Bias due to deviations from intended interventions		
If your aim for this study is to assess the effect of assignment to intervention, answer questions 4.1 and 4.2		
4.1. Were there deviations from the intended intervention beyond what would be expected in usual practice?	Details of 'intended' intervention unclear; retrospective analysis which considered patients with juvenile MLD who had received haematopoietic stem cell therapy (conditioning regimen, donor source and number of transplants varied between included patients).	NI
4.2. If Y/PY to 4.1: Were these deviations from intended intervention unbalanced between groups <i>and</i> likely to have affected the outcome?		NA
If your aim for this study is to assess the effect of starting and adhering to intervention, answer questions 4.3 to 4.6		
4.3. Were important co-interventions balanced across intervention groups?		NA
4.4. Was the intervention implemented successfully for most participants?		NA
4.5. Did study participants adhere to the assigned intervention regimen?		NA
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
Risk of bias judgement		NI

Bias due to missing data		
5.1 Were outcome data available for all, or nearly all, participants?	Four transplant-related deaths (treated patients) excluded. All non-transplanted patients survived for at least 5 years and there were 11 MLD progression-related deaths of un-treated patients during the 'observation period'; all 41 untreated patients were included in the analyses of gross motor function and cognitive function.	N

5.2 Were participants excluded due to missing data on intervention status?		N
5.3 Were participants excluded due to missing data on other variables needed for the analysis?	Brain imaging comparison included 36 non-transplanted patients and 20 transplanted patients.	Y
5.4 If PNN 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?		N
5.5 If PNN 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?		N
Risk of bias judgement		Serious

Bias in measurement of outcomes		
6.1 Could the outcome measure have been influenced by knowledge of the intervention received?	Some functional outcome measures involve subjective judgement and it was not clear whether assessors were aware of treatment history.	NI
6.2 Were outcome assessors aware of the intervention received by study participants?	MRI scans were scored by one experienced rater, using a validated severity scale and without knowledge of clinical status or outcome.	NI
6.3 Were the methods of outcome assessment comparable across intervention groups?	There was a lack of detailed reporting, particularly with respect to the un-treated, control group.	NI
6.4 Were any systematic errors in measurement of the outcome related to intervention received?		NI
Risk of bias judgement		NI

Bias in selection of the reported result		
Is the reported effect estimate likely to be selected, on the basis of the results, from...	Insufficient detail reported.	
7.1. ... multiple outcome <i>measurements</i> within the outcome domain?		NI
7.2 ... multiple <i>analyses</i> of the intervention-outcome relationship?		NI
7.3 ... different <i>subgroups</i> ?		PN
Risk of bias judgement		NI

Overall bias

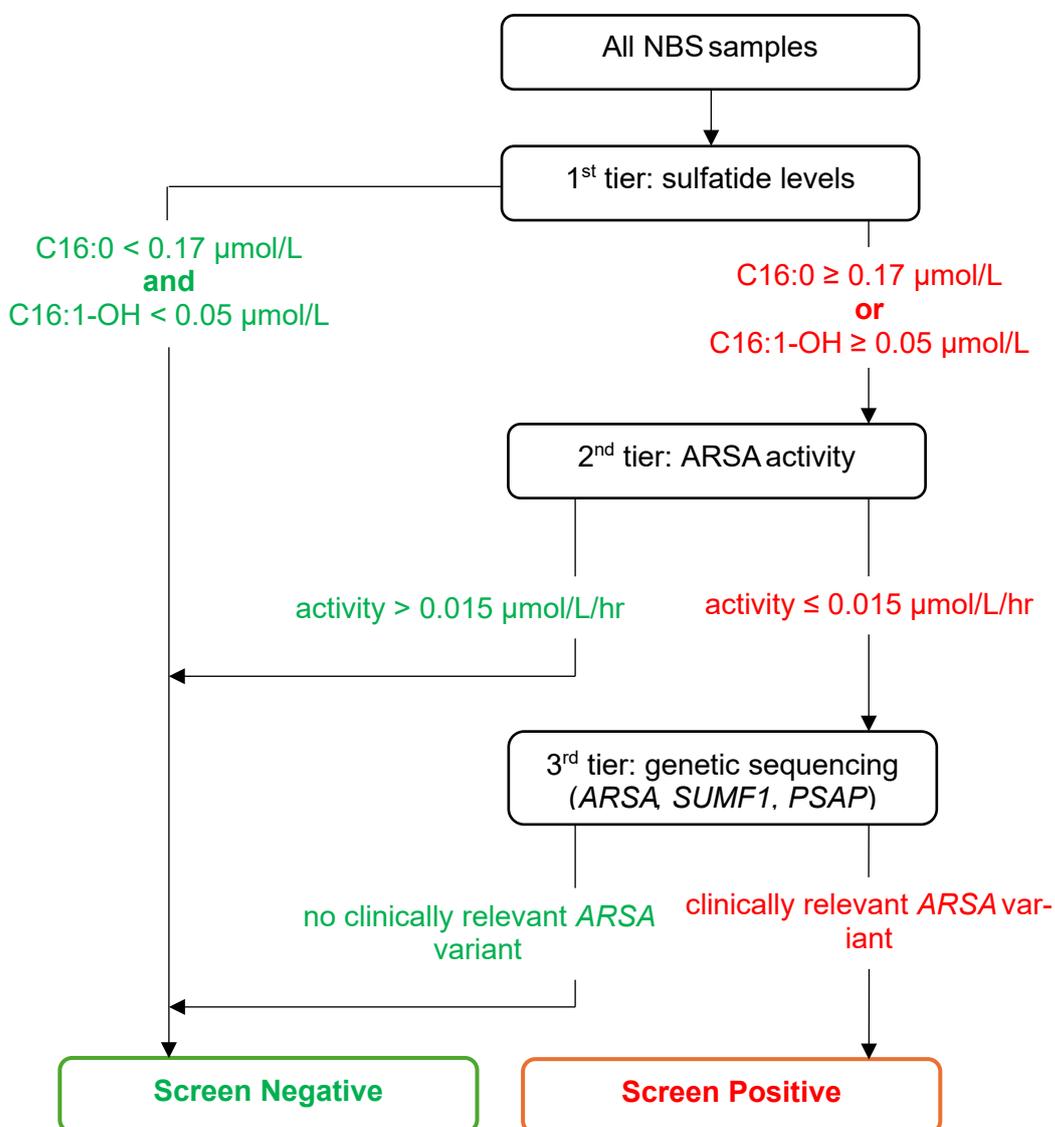
Risk of bias judgement	Absence of any consideration of confounding and serious issues in relation to selection of participants and missing data.	Critical
N: no; NA: not applicable; NI: no information; PN: probably no; PY: probably yes; Y: yes		

Appendix 4 - Screening algorithms evaluated by included studies

The following pages summarise the screening algorithms evaluated by studies included in this evidence summary, for NBS screening for MLD; algorithms are listed under the country where the evaluation was conducted.

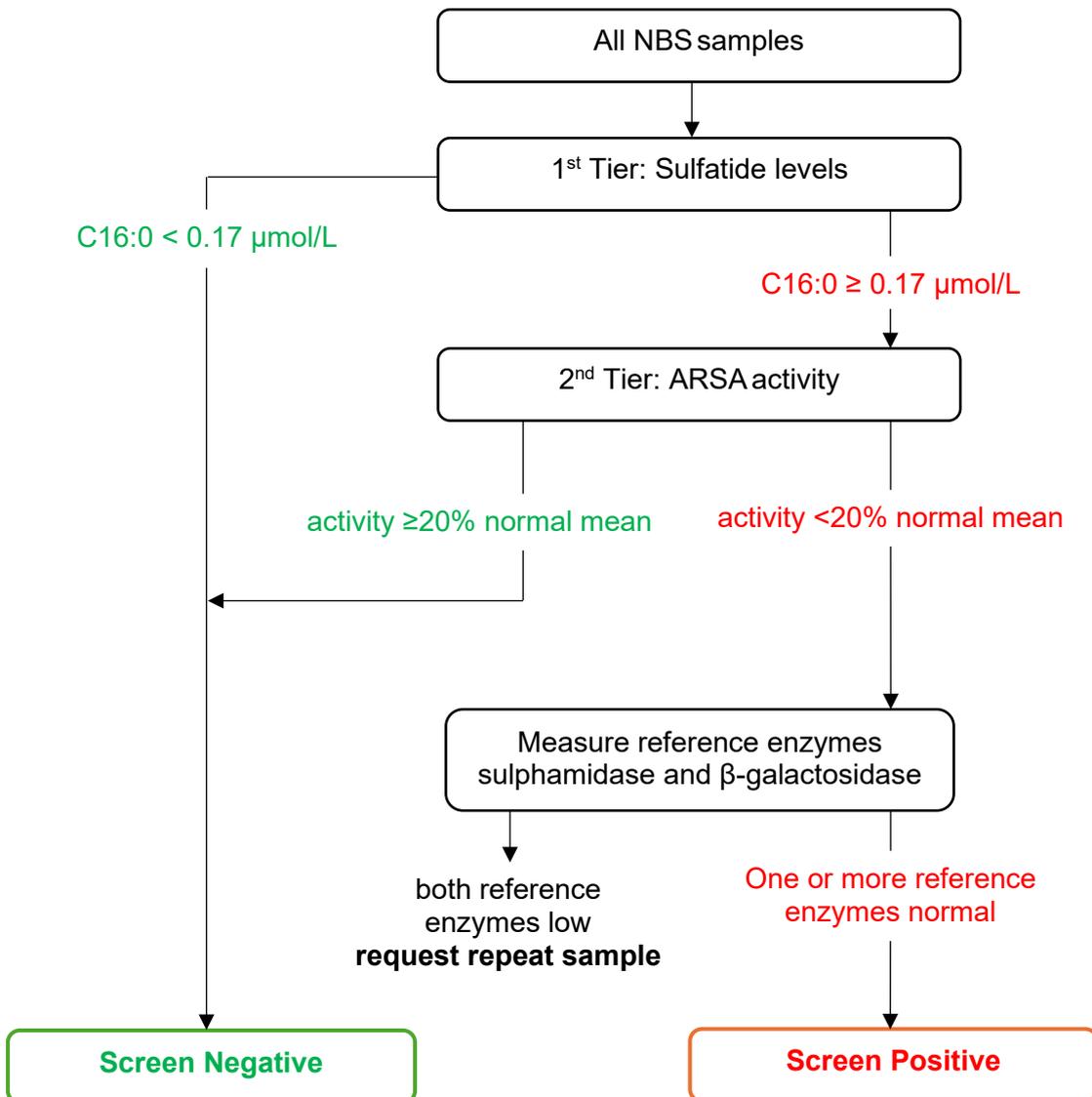
Germany

Laugwitz et al. (2024)²⁷ reported a pilot study evaluating a 3-tier strategy for NBS screening for MLD, comprising the quantification of C16:0-sulfatide and C16:1-OH-sulfatide using UPLC-MS/MS and measurement of ARSA enzymatic activity assay using tandem mass spectrometry, and where genetic sequencing of the DBS sample was classified as the 3rd tier of screening. The study used 109, 259 DBS obtained from newborns as part of the existing German newborn screening programme.



UK

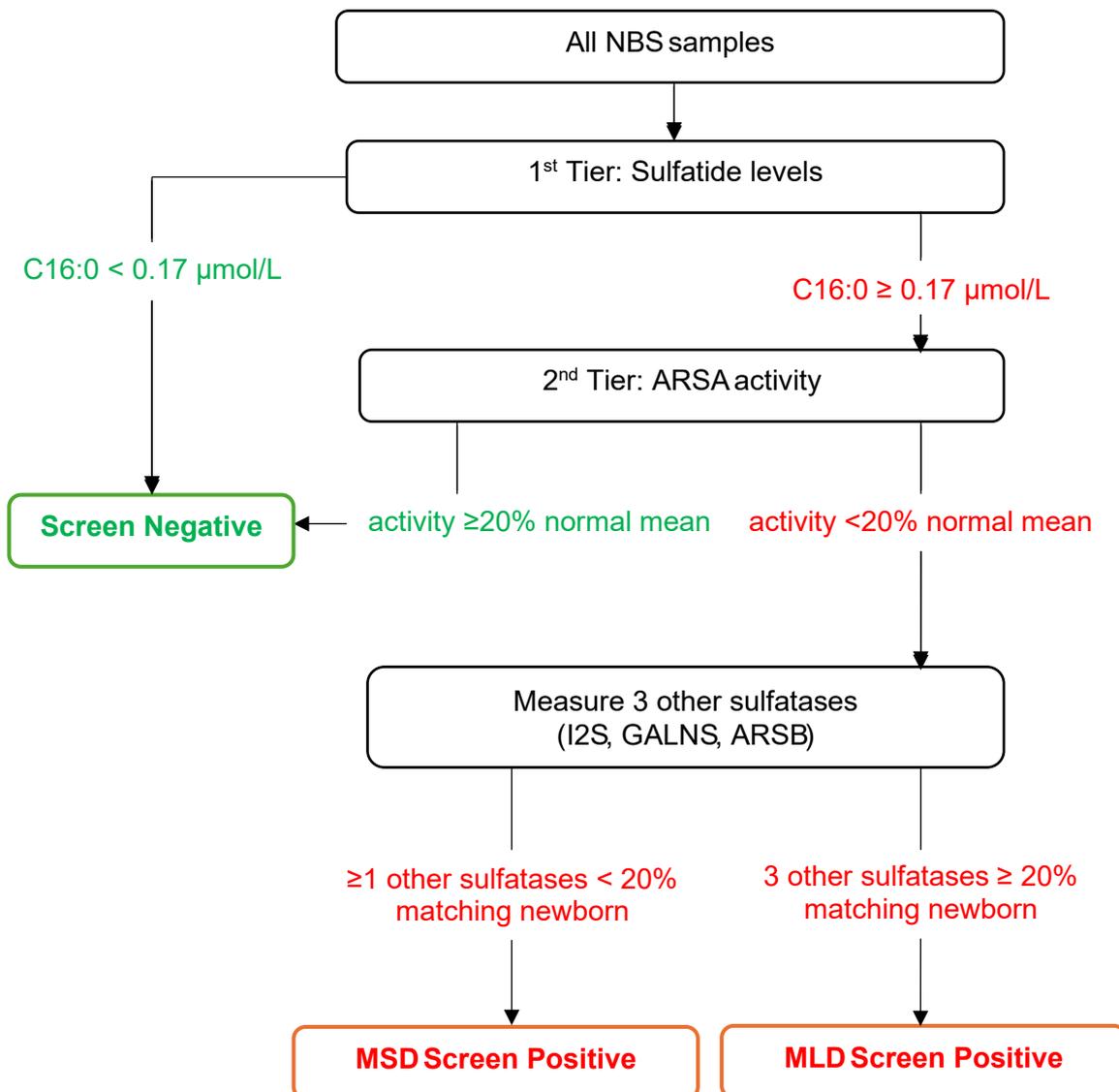
Wu et al. (2024)²⁸ reported a ‘pre-pilot’ study, conducted at the Royal Manchester Children’s Hospital, evaluating a 2-tier strategy for NBS screening for MLD, comprising the quantification of C16:0-sulfatide in using UPLC-MS/MS and measurement of ARSA enzymatic activity assay using tandem mass spectrometry. The study used 3,687 de-identified residual DBS samples from the Manchester Newborn Screening Laboratory.



US (Washington State)

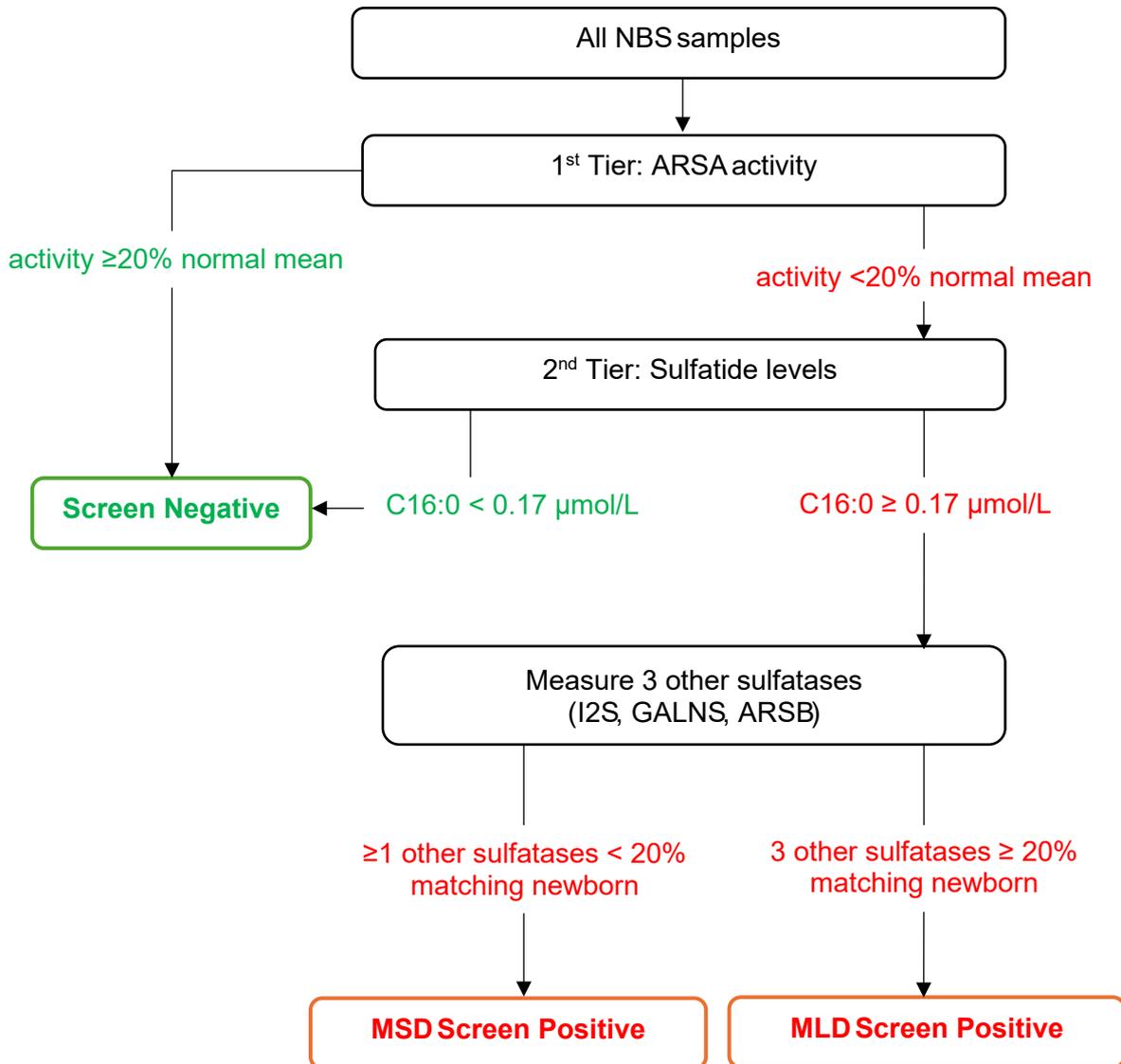
Hong et al. (2021)¹¹ assessed the feasibility of using a 2-tier strategy for NBS screening for MLD, comprising the quantification of C16:0-sulfatide in using UPLC-MS/MS and measurement of ARSA enzymatic activity assay using tandem mass spectrometry. The study used 27,355 de-identified DBS, shared by the Washington State Department of Health.

a)



Hong et al. (2021)¹¹ also conducted a preliminary exploration of a 2-tier strategy for NBS screening for MLD, where ARSA enzymatic activity was used as the 1st tier test (algorithm b), using 2,287 de-identified DBS.

b)



Appendix 5 - Published clinical guidelines on the management of MLD

NICE guidance HST18: Atidarsagene autotemcel for treating metachromatic leukodystrophy

HST18 recommends ARSA-cel (Libmeldy®), within its marketing authorisation, as an option for treating MLD with mutations in the Arylsulfatase A (ARSA) gene:

- for children who have late infantile or early juvenile types, with no clinical signs or symptoms
- for children who have the early juvenile type, with early clinical signs or symptoms, and who can still walk independently and have no cognitive decline

HST18 further notes that ARSA-cel should be delivered in a highly specialised service by a specialist multidisciplinary team.⁵

Newborn screening in MLD – European consensus-based recommendations on clinical management

The stated aim of this study, published in 2024, was to establish consensus among international experts in MLD and patient advocates on clinical management for cases of MLD identified through newborn screening.⁴³ It may therefore be argued that this guideline assumes that newborn screening for MLD is being or should be implemented. The concluding statement in this publication noted that, despite identified uncertainties and challenges, experts unanimously supported the implementation of newborn screening programmes for MLD, further stating that this endorsement was driven by the recognised efficacy of pre-symptomatic treatment and the technical feasibility of screening.⁴³ With respect to future work, the authors noted the need for harmonised management and integration of national screening programmes, structured data collection and monitoring of screening programmes for evidence generation and to inform future guideline development, and involvement of patient representatives in the development of recommendations.⁴³

The guideline development process used a real-time Delphi procedure, involving a multidisciplinary expert panel (n=22), including paediatric and adult neurologists, physicians with expertise in paediatric and adult inherited metabolic diseases, paediatric and adult haematologists, paediatricians and a geneticist, from 11 countries (Canada, Denmark, France, Germany, Israel, Italy, the Netherlands, Norway, Sweden, UK, US). Questions were based on findings from a literature review (details not reported) and workshops. Responses required participants to indicate agreement or disagreement on a 3-point Likert scale, select preferences in single/multiple choice questions, or provide open-ended answers. Recommendations were rated according to level of consensus: Level A 100%; Level B 75-99%; Level C 50-74% or >75% but >25% neutral votes. No grading of recommendations, based on level of evidence, was reported.⁴³

The guideline included 57 recommendations under the headings ‘communication and counselling’ (n=13); ‘confirmatory diagnostics’ (n=15); ‘preservation of biomaterial’

UKNSC external review — Newborn screening for metachromatic leukodystrophy (MLD), November 2025

(n=1); 'prediction of symptom onset' (n=10); 'definition of disease onset' (n=2); 'treatment' (n=14); 'monitoring' (n=1); 'newborn screening for MLD' (n=1).⁴³ The following is a list of those recommendations which were classified as Level A and which also had no associated neutral votes:⁴³

Newborn screening for MLD

'Newborn screening for MLD is recommended and aligns with established criteria'

Monitoring

'Regular post-treatment follow-up in an expert centre is recommended in all newborn screening identified patients.'

Treatment

'It is strongly recommended to treat MLD patients before they exhibit MLD-related symptoms, late infantile MLD.'

'It is strongly recommended that late infantile patients are treated with autologous haematopoietic stem cell gene therapy (ARSA-cel).'

'It is recommended to treat pre-symptomatic early juvenile patients with allogenic haematopoietic stem cell transplantation only in case ARSA-cel is not available, late juvenile MLD.'

'It is recommended to schedule early juvenile patients for apheresis between 9 and 12 months (>8 kg body weight).'

'It is not recommended to schedule adult patients for allogenic haematopoietic stem cell transplantation at a predefined age, but to be guided by a case-to-case decision of the treatment eligibility panel.'

Communication and counselling

'It is strongly recommended that the family is informed about the contact at an expert centre when a positive screening result is communicated.'

'It is strongly recommended to arrange a treatment eligibility panel discussion according to the procedure from the European Reference Network on Rare Neurological Diseases (ERN-RND) and the MLD initiative to discuss treatment eligibility.'

Confirmatory diagnostics

'It is strongly recommended to offer comprehensive genetic counselling to identify potentially affected relatives.'

Preservation of biomaterial

'It is strongly recommended to archive bio samples collected in newborn screening identified cases to enable future studies according to local ethics votes.'

Prediction of symptom onset

‘Late onset can be predicted for individuals harbouring a known genotype with late juvenile or adult onset well reported in literature.’

Appendix 6 - UKNSC reporting checklist for evidence summaries

All items on the UKNSC 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 39.

Table 39: Reporting checklist

Section	Item	Page no.
Title and summaries		
Title sheet	Identify the review as a UKNSC Evidence summary	Title page
Plain English summary	Plain English description of the executive summary	6 to 7
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	8 to 12
Introduction and Approach		

Section	Item	Page no.
Background and objectives	<p>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</p> <p>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.</p> <p>Method – briefly outline the rapid review methods used.</p>	13 to 18
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	19 to 22
Appraisal for quality/ risk of bias tool	Details of tool/checklist used to assess quality, e.g. QUADAS2, CASP, SIGN, AMSTAR.	23
Search strategy and study selection		
Databases/ sources searched	Give details of all databases searched (including platform/ interface and coverage dates) and date of final search.	23 to 25

Section	Item	Page no.
Search strategy and results	<p>Present the full search strategy for at least one database(usually a version of MEDLINE), including limits and search filters if used.</p> <p>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.</p>	62 to 76 (Appendix 1)
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.	19
Study level reporting of results (for each key question)		

Section	Item	Page no.
Study level reporting, results and risk of bias assessment	For each study, produce a table that includes the full citation and a summary of the data relevant to the question (for example, study size, PICO, follow-up period, outcomes reported, statistical analyses etc.).	Study level reporting: 28 to 41
	Provide a simple summary of key measures, effect estimates and confidence intervals for each study where available.	(Tables 3 to 6)
	For each study, present the results of any assessment of quality/risk of bias.	42 to 47
		(Tables 7 to 12)
		56 to 58
		(Table 13)
		Quality assessment: 84 to 103 (Appendix 3)

Section	Item	Page no.
Additional analyses	Describe additional analyses (for example, sensitivity, specificity, PPV, etc.) carried out by the reviewer.	<p>Where possible, incidence of screen-detected MLD was calculated</p> <p>37 to 38 (Table 4)</p> <p>Where possible, PPV (assuming no FN) was estimated</p> <p>30 and 32</p>

Question level synthesis

Section	Item	Page no.
Description of the evidence	For each question, give numbers of studies screened, assessed for eligibility, and inclusion in the review, with summary reasons for exclusion	<p>26 to 27 and 77 to 83 (Appendix 2)</p> <p>Searches were conducted for the whole evidence summary and not separately by question.</p> <p>An overall PRISMA flow chart and details of included and excluded studies are provided in Appendix 2</p>
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 compartments should inform the reviewer’s judgement on whether the criterion is “met”, “not met” or “uncertain”: quantity; quality; applicability and consistency.	<p>40 to 41</p> <p>45 to 46</p> <p>58 to 59</p>

Section	Item	Page no.
Summary of findings	Provide a description of the evidence reviewed and included for each question, with reference to their eligibility for inclusion.	41
		46 to 47
	Summarise the main findings including the quality/ risk of bias issues for each question.	59
	Have the criteria addressed been “met”, “not met” or “uncertain”?	
Review Summary		
Conclusions and implications for policy	Do findings indicate whether screening should be recommended?	60 to 61
	IS further work warranted?	
	Are there gaps in the evidence highlighted by the review?	
Limitations	Discuss limitations of the available evidence and of the review methodology if relevant.	61

References

1. Gomez-Ospina N. Arylsulfatase A deficiency. In: Pagon RA, Wallace SE, Bean LJH, Gripp KW, Amemiya A, editors. GeneReviews®. Seattle (WA): University of Washington; 2024 [accessed 2.7.25].
2. Shaimardanova AA, Chulpanova DS, Solovyeva VV, Mullagulova AI, Kitaeva KV, Allegrucci C, et al. Metachromatic leukodystrophy: diagnosis, modeling, and treatment approaches. *Front Med (Lausanne)*. 2020;7:Article 576221.
3. Great Ormond Street Hospital for Children, NHS Foundation Trust. Metachromatic leukodystrophy (late infantile form) [Internet]. n.d. [accessed 2.7.2025] [
4. National Institute of Neurological Disorders and Stroke. Metachromatic leukodystrophy [Internet] n.d. [accessed 2.7.25] [Available from: <https://www.ninds.nih.gov/health-information/disorders/metachromatic-leukodystrophy#:~:text=The%20prognosis%20for%20MLD%20is,years%20following%20onset%20of%20symptoms>].
5. National Institute for Health and Care Excellence. Atidarsagene autotemcel for treating metachromatic leukodystrophy. NICE Highly specialised technologies guidance (HST18) [Internet]. London: NICE; 2022 [accessed 2.7.25].
6. Mayo Clinic. Metachromatic leukodystrophy [Internet] 2020 [accessed 2.7.25] [Available from: <https://www.mayoclinic.org/diseases-conditions/metachromatic-leukodystrophy/symptoms-causes/syc-20354733>].
7. UK National Screening Committee, Department of Health & Social Care. Specification document. Evidence summary: newborn screening for metachromatic leukodystrophy (MLD). Department of Health & Social Care; 2024 [accessed 9.7.24].
8. Spacil Z, Babu Kumar A, Liao HC, Auray-Blais C, Stark S, Suhr TR, et al. Sulfatide analysis by mass spectrometry for screening of metachromatic leukodystrophy in dried blood and urine samples. *Clin Chem*. 2016;62(1):279-86.
9. Tan MA, Dean CJ, Hopwood JJ, Meikle PJ. Diagnosis of metachromatic leukodystrophy by immune quantification of arylsulphatase A protein and activity in dried blood spots. *Clin Chem*. 2008;54(11):1925-7.
10. Hong X, Kumar AB, Daiker J, Yi F, Sadilek M, De Mattia F, et al. Leukocyte and dried blood spot arylsulfatase A assay by tandem mass spectrometry. *Anal Chem*. 2020;92(9):6341-8.
11. Hong X, Daiker J, Sadilek M, Ruiz-Schultz N, Kumar AB, Norcross S, et al. Toward newborn screening of metachromatic leukodystrophy: results from analysis of over 27,000 newborn dried blood spots. *Genet Med*. 2021;23(3):555-61.
12. Wasserstein M, Caggana M, Gelb MH, Goldenberg A, Kelly N, Matern D, et al. ScreenPlus: a comprehensive, dynamic, multi-disorder newborn screening pilot program. *Mol Genet Metab*. 2020;129(2):S160.
13. Bekri S, Bley A, Brown HA, Chanson C, Church HJ, Gelb MH, et al. Higher precision, first tier newborn screening for metachromatic leukodystrophy using 16:1-OH-sulfatide. *Mol Genet Metab*. 2024;142(1).
14. Schlotawa L, Dierks T, Christoph S, Cloppenburg E, Ohlenbusch A, Korenke GC, et al. Severe neonatal multiple sulfatase deficiency presenting with hydrops fetalis in a preterm birth patient. *JIMD Rep*. 2019;49(1):48-52.
15. Schlotawa L, Adang L, De Castro M, Ahrens-Nicklas R. Multiple sulfatase deficiency. In: Adam MP, Feldman J, Mirzaa GM, et al., editors. GeneReviews®. Seattle (WA): University of Washington; 2019 [accessed 2.7.25].

16. Laugwitz L, Santhanakumaran V, Spieker M, Boehringer J, Bender B, Gieselmann V, et al. Extremely low arylsulfatase A enzyme activity does not necessarily cause symptoms: a long-term follow-up and review of the literature. *JIMD Rep.* 2022;63(4):292-302.
17. Ridsdale R, Kroll C, Sanders K, Olgesbee D, Rinaldo P, Hopwood J, et al. Newborn screening (NBS) for metachromatic leukodystrophy (MLD): results from a study of 100,000 deidentified NBS samples. *Mol Genet Metab.* 2017;120(1-2):S115.
18. Umapathysivam K, Whittle AM, Ranieri E, Bindloss C, Ravenscroft EM, van Diggelen OP, et al. Determination of acid alpha-glucosidase protein: evaluation as a screening marker for Pompe disease and other lysosomal storage disorders. *Clin Chem.* 2000;46(9):1318-25.
19. Blenda AV. Medscape: Metachromatic leukodystrophy treatment & management [Internet] 2024 [accessed 2.7.25] [Available from: <https://emedicine.medscape.com/article/951840-treatment>]
20. Armstrong N, Olaye A, Noake C, Pang F. A systematic review of clinical effectiveness and safety for historical and current treatment options for metachromatic leukodystrophy in children, including atidarsagene autotemcel. *Orphanet J Rare Dis.* 2023;18(1):248.
21. Genomics Education Programme. A ground-breaking new gene therapy has saved the life of its very first NHS patient [Internet]. NHS England; 2023 [accessed 2.7.25].
22. Manchester University NHS Foundation Trust. First baby receives life-saving gene therapy on NHS at Royal Manchester Children's Hospital [News posted 15.02.2023] [Internet] 2024 [accessed 4.9.24] [Available from: <https://mft.nhs.uk/2023/02/15/first-baby-receives-life-saving-gene-therapy-on-nhs-at-royal-manchester-childrens-hospital/>].
23. U.S. Food and Drug Administration. FDA news release: FDA approves first gene therapy for children with metachromatic leukodystrophy [Internet]: FDA; 2024 [accessed 2.7.25] [Available from: <https://www.fda.gov/news-events/press-announcements/fda-approves-first-gene-therapy-children-metachromatic-leukodystrophy>].
24. Health Resources and Services Administration (HRSA). Recommended uniform screening panel core conditions [Internet] Rockville, MD: HRSA; 2024 [accessed 20.8.24].
25. Health Resources and Services Administration (HRSA). Summary of nominated conditions to the recommended uniform screening panel (RUSP) [Internet]. Rockville, MD: HRSA; 2023 [accessed 20.8.24].
26. UK National Screening Committee. Guidance - Criteria for a population screening programme [Internet] 2022 [accessed 20.8.24] [Available from: <https://www.gov.uk/government/publications/evidence-review-criteria-national-screening-programmes/criteria-for-appraising-the-viability-effectiveness-and-appropriateness-of-a-screening-programme>].
27. Laugwitz L, Mechtler TP, Janzen N, Oliva P, Kasper A-R, Teunissen CE, et al. Newborn screening and presymptomatic treatment of metachromatic leukodystrophy. *N Engl J Med.* 2024;391(13):1256-8.
28. Wu THY, Brown HA, Church HJ, Kershaw CJ, Hutton R, Egerton C, et al. Improving newborn screening test performance for metachromatic leukodystrophy: recommendation from a pre-pilot study that identified a late-infantile case for treatment. *Mol Genet Metab.* 2024;142(1):108349.
29. Fumagalli F, Calbi V, Natali Sora MG, Sessa M, Baldoli C, Rancoita PMV, et al. Lentiviral haematopoietic stem-cell gene therapy for early-onset metachromatic leukodystrophy: long-term results from a non-randomised, open-label, phase 1/2 trial and expanded access. *Lancet.* 2022;399(10322):372-83.
30. Groeschel S, Kuhl JS, Bley AE, Kehrer C, Weschke B, Doring M, et al. Long-term outcome of allogeneic hematopoietic stem cell transplantation in patients with juvenile metachromatic leukodystrophy compared with nontransplanted control patients. *JAMA Neurol.* 2016;73(9):1133-40.

31. Bean K, Jones SA, Chakrapani A, Vijay S, Wu T, Church H, et al. Exploring the cost-effectiveness of newborn screening for metachromatic leukodystrophy (MLD) in the UK. *Int J Neonatal Screen*. 2024;10(3):45.
32. Centre for Reviews and Dissemination. *Systematic Reviews: CRD's guidance for undertaking reviews in health care* [Internet]. York: University of York; 2009 [accessed 2.7.25].
33. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. *Cochrane handbook for systematic reviews of interventions version 6.3 (updated February 2022)* [Internet]. Cochrane; 2022 [accessed 4.3.22].
34. Canadian Agency for Drugs and Technologies in Health. *PRESS - Peer Review of Electronic Search Strategies: 2015 Guideline Explanation and Elaboration (PRESS E&E)* [Internet]. Ottawa: CADTH; 2016 [accessed 2.7.25].
35. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-36.
36. Yang B, Mallett S, Takwoingi Y, Davenport CF, Hyde CJ, Whiting PF, et al. QUADAS-C: a tool for assessing risk of bias in comparative diagnostic accuracy studies. *Ann Intern Med*. 2021;174(11):1592-9.
37. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. 2016;355:i4919.
38. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. *BMJ*. 1996;313(7052):275-83.
39. Deeks JJ, Bossuyt PM, Leeflang MM, Takwoingi Y, editors. *Cochrane handbook for systematic reviews of diagnostic test accuracy. Version 2.0 (updated July 2023)* [Internet]. Cochrane; 2023 [accessed 2.7.25].
40. Sessa M, Lorioli L, Fumagalli F, Acquati S, Redaelli D, Baldoli C, et al. Lentiviral haemopoietic stem-cell gene therapy in early-onset metachromatic leukodystrophy: an ad-hoc analysis of a non-randomised, open-label, phase 1/2 trial. *Lancet*. 2016;388(10043):476-87.
41. Øberg A. [New changes in newborn screening [Internet]]. Oslo: Oslo University Hospital (OUS); 2025 [accessed 4.2.25].
42. Charles River Associates. *A landscape assessment of newborn screening (NBS) in Europe: executive summary* [Internet]. 2024 [accessed 18.2.25].
43. Laugwitz L, Schoenmakers DH, Adang LA, Beck-Woedl S, Bergner C, Bernard G, et al. Newborn screening in metachromatic leukodystrophy - European consensus-based recommendations on clinical management. *Eur J Paediatr Neurol*. 2024;49:141-54.
44. Schwarz M, Oliva P, Mechtler T, Streubel B, Chanson C, Essing MM, et al. Newborn screening for metachromatic leukodystrophy in Northern Germany - a prospective study [Poster]. 17th WORLDSymposium; Virtual2021.
45. Schwartz O, Vill K, Pfaffenlehner M, Behrens M, Weiß C, Johannsen J, et al. Clinical Effectiveness of Newborn Screening for Spinal Muscular Atrophy: A Nonrandomized Controlled Trial. *JAMA Pediatr*. 2024;178(6):540-47.
46. NCT. Newborn Screening for Spinal Muscular Atrophy (SMA-NBS). <https://clinicaltrials.gov/study/NCT05481164>. 2024.
47. Biffi A, Montini E, Lorioli L, Cesani M, Fumagalli F, Plati T, et al. Lentiviral hematopoietic stem cell gene therapy benefits metachromatic leukodystrophy. *Science*. 2013;341(6148):1233-158.
48. Calbi V, Fumagalli F, Sessa M, Zambon A, Baldoli C, Cugnata F, et al. Lentiviral hematopoietic stem and progenitor cell gene therapy (HSPC-GT) for metachromatic

leukodystrophy (MLD): clinical outcomes from 33 patients. *Bone Marrow Transplant.* 2020;55:72-3.

49. Calbi V, Fumagalli F, Gallo V, Recupero S, De Mattia F, Fratini ES, et al. Lentiviral haematopoietic stem cell gene therapy for metachromatic leukodystrophy: results in 5 patients treated under nominal compassionate use. *Mol Genet Metab.* 2023;138(2).
50. Essing M, Zambon A, Gallo V, Baldoli C, Cugnata F, Rancoita PMV, et al. Lentiviral hematopoietic stem and progenitor cell gene therapy for metachromatic leukodystrophy (MLD): clinical outcomes from 38 patients. *Neuropediatrics.* 2021;52(Suppl 1).
51. Fumagalli F, Calbi V, Zambon A, Ciotti F, Lorioli L, Sessa M, et al. Update on safety and efficacy of lentiviral hematopoietic stem cell gene therapy (HSC-GT) for metachromatic leukodystrophy (MLD). *Eur J Paediatr Neurol.* 2017;21(Suppl 1):e20.
52. Fumagalli F, Calbi V, Sessa M, Cugnata F, Rancoita PMV, Acquati S, et al. Lentiviral hematopoietic stem cell gene therapy (HSCGT) for metachromatic leukodystrophy (MLD) provides sustained clinical benefit. *Bone Marrow Transplant.* 2019;54:76-7.
53. Fumagalli F, Calbi V, Sessa M, Zambon A, Baldoli C, Cugnata F, et al. Lentiviral (LV) hematopoietic stem cell gene therapy (HSC-GT) for metachromatic leukodystrophy (MLD) provides sustained clinical benefit. *Ann Neurol.* 2019;86(Suppl 23):S163-S4.
54. Fumagalli F, Calbi V, Sessa M, Zambon A, Baldoli C, Cugnata F, et al. Lentiviral (LV) hematopoietic stem cell gene therapy (HSC-GT) for metachromatic leukodystrophy (MLD) provides sustained clinical benefit. *J Inherit Metab Dis.* 2019;42(Suppl 1):7-8.
55. Fumagalli F, Calbi V, Sessa M, Zambon A, Baldoli C, Rancoita PMV, et al. Lentiviral hematopoietic stem and progenitor cell gene therapy (HSPC-GT) for metachromatic leukodystrophy (MLD): clinical outcomes from 33 patients. *Mol Genet Metab.* 2020;129(2):S59.
56. Fumagalli F, Calbi V, Zambon A, Gallo V, Baldoli C, Cugnata F, et al. Lentiviral haematopoietic stem and progenitor cell gene therapy for metachromatic leukodystrophy (MLD): clinical outcomes from 38 Patients. *Dev Med Child Neurol.* 2022;64(Suppl 1):55-6.
57. Fumagalli F, Calbi V, De Mattia F, Zambon AA, Gallo V, Recupero S, et al. Long-term clinical outcomes of atidarsagene autotemcel (autologous hematopoietic stem cell gene therapy [HSC-GT] for metachromatic leukodystrophy) with up to 11 years follow-up. *Mol Genet Metab.* 2023;138(2).
58. Kehrer C, Groeschel S, Doering M, Krageloh-Mann I. 5-year follow-up in hematopoietic stem cell transplantation in 2 patients with late-infantile metachromatic leukodystrophy in comparison to an untreated cohort. *Eur J Paediatr Neurol.* 2013;17(Suppl 1):S132.
59. Martin HR, Poe MD, Provenzale JM, Kurtzberg J, Mendizabal A, Escolar ML. Neurodevelopmental outcomes of umbilical cord blood transplantation in metachromatic leukodystrophy. *Biol Blood Marrow Transplant.* 2013;19(4):616-24.
60. Orchard PJ, Gupta A, Braun J, Adang LA, Pillai NR, Lund T, et al. Compassionate use of OTL-200 for patients with metachromatic leukodystrophy. *Mol Genet Metab.* 2023;138(2).
61. Sevin C, Roujeau T, Cartier N, Baugnon T, Adamsbaum C, Piraud M, et al. Intracerebral gene therapy in children with metachromatic leukodystrophy: results of a phase I/II trial. *Mol Genet Metab.* 2018;123(2):S129.
62. van Rappard DF, Boelens JJ, van Egmond ME, Kuball J, van Hasselt PM, Oostrom KJ, et al. Efficacy of hematopoietic cell transplantation in metachromatic leukodystrophy: the Dutch experience. *Blood.* 2016;127(24):3098-101.
63. Yoon IC, Bascou NA, Poe MD, Escolar ML. Long-term neurodevelopmental, neurophysiological, and neuroradiological outcomes of hematopoietic stem cell transplantation for treatment of late-infantile metachromatic leukodystrophy. *Mol Genet Metab.* 2020;129(2):S164-S5.

64. National Institute for Health and Care Excellence. OTL-200 for treating metachromatic leukodystrophy [ID1666]. Highly Specialised Technology evaluation: Committee papers [Internet]. London: NICE; 2021 [accessed 19.2.25].
65. Fumagalli F, Calbi V, Gallo V, Zambon AA, Recupero S, Ciotti F, et al. Long-Term Effects of Atidarsagene Autotemcel for Metachromatic Leukodystrophy. *N Engl J Med*. 2025;392:1609-20.
66. UK National Screening Committee. Evaluation of NHS newborn screening for SCID extended to March 2024. Professor Anne Mackie, 13 June 2023 [Blog] [Internet] 2023 [accessed 12.11.2025] [Available from: <https://nationalscreening.blog.gov.uk/2023/06/13/evaluation-of-nhs-newborn-screening-for-scid-extended-to-march-2024/>].
67. Biffi A, Cesani M, Fumagalli F, Del Carro U, Baldoli C, Canale S, et al. Metachromatic leukodystrophy - mutation analysis provides further evidence of genotype-phenotype correlation. *Clin Genet*. 2008;74(4):349-57.
68. Bean K, Jones S, Chakrapani A, Vijay S, Wu THY, Church HJ, et al. POSB95 The cost-effectiveness of newborn screening for metachromatic leukodystrophy (MLD) in the UK. *Value Health*. 2022;25(1 Suppl):S79.
69. Pang F, Dean R, Jensen I, Tehard B, Roze S, Olaye A, et al. The cost-effectiveness of atidarsagene autotemcel for the treatment of metachromatic leukodystrophy in France. *Mol Genet Metab*. 2022;135(2):S93.
70. Pang F, Dean R, Jensen I, Tehard B, Roze S, Olaye A, et al. The cost-effectiveness of atidarsagene autotemcel for the treatment of metachromatic leukodystrophy (MLD) in France. *Value Health*. 2022;25(7 Suppl):S339.
71. Pang F, Dean R, Jensen I, Olaye A, Miller B. PRO30 The cost-effectiveness of otl-200 for the treatment of metachromatic leukodystrophy (MLD). *Value Health*. 2021;24(Suppl 1):S203.
72. Pang F, Dean R, Jensen I, Bean K, Fields C, Miller B. The cost-effectiveness of OTL-200 for the treatment of metachromatic leukodystrophy (MLD) in the US. *Mol Genet Metab*. 2023;138(2).
73. Barcenas M, Suhr TR, Scott CR, Turecek F, Gelb MH. Quantification of sulfatides in dried blood and urine spots from metachromatic leukodystrophy patients by liquid chromatography/electrospray tandem mass spectrometry. *Clin Chim Acta*. 2014;433:39-43.
74. Boucher AA, Miller W, Shanley R, Ziegler R, Lund T, Raymond G, et al. Long-term outcomes after allogeneic hematopoietic stem cell transplantation for metachromatic leukodystrophy: the largest single-institution cohort report. *Orphanet J Rare Dis*. 2015;10:94.
75. Calbi V, Fumagalli F, Fratini ES, Palasciano A, Recupero S, Gallo V, et al. Blood sulfatides as disease biomarker for metachromatic leukodystrophy: disease characterization, early diagnosis, and response to treatment. *Mol Genet Metab*. 2023;138(2):107043.
76. Calbi V, Fumagalli F, Mattia FD, Zambon A, Gallo V, Recupero S, et al. Atidarsagene autotemcel (hematopoietic stem cell gene therapy) preserves cognitive and motor development in metachromatic leukodystrophy with up to 12 years follow-up. *Mol Ther*. 2024;32(4 Suppl 1):1.
77. Chang SC, Bergamasco A, Bonnin M, Arredondo-Bisono T, Moride Y. Birth prevalence of metachromatic leukodystrophy: a systematic literature review. *Mol Genet Metab*. 2023;138(2):107249.
78. Elmonem MA, Ramadan DI, Issac MS, Selim LA, Elkateb SM. Blood spot versus plasma chitotriosidase: a systematic clinical comparison. *Clin Biochem*. 2014;47(1-2):38-43.
79. Fahim SM, Lin G, Suh K, Carlson JJ, Richardson M, Herce-Hagiwara B, et al. Atidarsagene autotemcel for metachromatic leukodystrophy. *J Manag Care Spec Pharm*. 2024;30(2):201-5.

80. Ferraiuolo L, Kaspar BK. Gene delivery improvement for treating the lysosomal storage disorder metachromatic leukodystrophy. *Hum Gene Ther.* 2012;23(8):793-5.
81. Gelb MH, Wasserstein M, Orsini JJ, Oliva P, Mechtler TP, Brown H, et al. Newborn screening for metachromatic leukodystrophy (MLD): an overview of ongoing and future studies. *Mol Genet Metab.* 2023;138(2).
82. Horgan C, Watts K, Ram D, Rust S, Hutton R, Jones S, et al. A retrospective cohort study of Libmeldy (atidarsagene autotemcel) for MLD: what we have accomplished and what opportunities lie ahead. *JIMD Rep.* 2023;64(5):346-52.
83. Jones S, Davison J, Mooney P, Campbell L, Baldock L, Wallington M, et al. Demographic and clinical characteristics of patients with metachromatic leukodystrophy in the United Kingdom: interim results from an observational real-world study. *Mol Genet Metab.* 2021;132(2):S53.
84. Jones SA, Cheillan D, Chakrapani A, Church HJ, Heales S, Wu THY, et al. Application of a novel algorithm for expanding newborn screening for inherited metabolic disorders across Europe. *Int J Neonatal Screen.* 2022;8(1).
85. Kehrer C, Groschel S, Muller I, Krageloh-Mann I. Stem cell transplantation (SCT) in metachromatic leukodystrophy (MLD) - results from 9 patients. *Neuropediatrics.* 2012;43(2).
86. Morton G, Thomas S, Roberts P, Clark V, Imrie J, Morrison A. The importance of early diagnosis and views on newborn screening in metachromatic leukodystrophy: results of a caregiver survey in the UK and Republic of Ireland. *Orphanet J Rare Dis.* 2022;17(1):403.
87. Oliva P, Mechtler TP, Schwarz M, Streubel B, Chanson C, Janzen N, et al. A MLD newborn screening pilot-study for metachromatic leukodystrophy in Germany: results of the first 12 months. *Mol Genet Metab.* 2023;138(2):107254.
88. Pettazzoni M, Rotaru I, Ruet S, Cheillan D, Sevin C, Acquaviva-Bourdain C, et al. LC-MS/MS quantification of three C16 sulfatide species in dried blood spots for the diagnosis and treatment monitoring of metachromatic leukodystrophy. *Mol Genet Metab.* 2023;138(2):107265.
89. Ruiz-Schultz N, Sant D, Norcross S, Dansithong W, Hart K, Asay B, et al. Methods and feasibility study for exome sequencing as a universal second-tier test in newborn screening. *Genet Med.* 2021;23(4):767-76.
90. Suhr D. Measuring sulfatide in blood enables newborn screening for metachromatic leukodystrophy (MLD). *Mol Genet Metab.* 2017;120(1-2):S128.
91. Van Rappard DF, Boelens JJ, Pouwels PJW, Hollak CEM, Van Der Knaap MS, Wolf NI. The effectiveness of hematopoietic cell transplantation (HCT) in metachromatic leukodystrophy (MLD): promising results. *Eur J Paediatr Neurol.* 2015;19(Suppl 1):S1.
92. Wiesinger T, Schwarz M, Oliva P, Mechtler TP, Streubel B, Chanson C, et al. Newborn screening for metachromatic leukodystrophy in Northern Germany - a prospective study. *Mol Genet Metab.* 2021;132(2):S112.
93. Fumagalli F, Zambon AA, Rancoita PMV, Baldoli C, Canale S, Spiga I, et al. Metachromatic leukodystrophy: A single-center longitudinal study of 45 patients. *J Inherit Metab Dis.* 2021;44(5):1151-64.