NIHR | National Institute for Health and Care Research

Health Technology Assessment

Check for updates

Extended Research Article

Evaluating whole genome sequencing for rare diseases in newborn screening: evidence synthesis from a series of systematic reviews

Karoline Freeman,[1*] Jacqueline Dinnes,[2] Bethany Shinkins,[1] Corinna Clark,[1] Inès Kander,[1] Katie Scandrett,[2] Shivashri Chockalingam,[1] Aziza Osman,[1] Naila Dracup,[1] Rachel Court,[1] Furqan Butt,[1] Cristina Visintin,[3] James R Bonham,[4] David Elliman,[5] Graham Shortland,[3] Anne Mackie,[3] Zosia Miedzybrodzka,[6] Sian M Morgan,[7] Felicity Boardman,[1] Yemisi Takwoingi[2] and Sian Taylor-Phillips[1]

[1]Warwick Medical School, University of Warwick, Coventry, UK
[2]Institute of Applied Health Research, University of Birmingham, Birmingham, UK
[3]UK National Screening Committee, London, UK
[4]Sheffield Children's NHS Foundation Trust, Sheffield, UK
[5]Great Ormond Street Hospital, London, UK
[6]Institute of Medical Sciences, University of Aberdeen, Aberdeen, UK
[7]All Wales Genetics Laboratory, Institute of Medical Genetics, University Hospital of Wales, Cardiff, UK

[*]**Corresponding author** K.Freeman@warwick.ac.uk

Published December 2025
DOI: 10.3310/DJRF1124

Scientific summary

Evaluating whole genome sequencing for rare diseases in newborn screening: evidence synthesis from a series of systematic reviews

Health Technology Assessment 2025; Vol. 29: No. 65
DOI: 10.3310/DJRF1124

NIHR Journals Library www.journalslibrary.nihr.ac.uk

Scientific summary

Background

In 2021, Genomics England Limited (GEL) launched its Generation Study of whole genome sequencing (WGS) to screen for over 200 rare diseases in 100,000 newborns to explore its potential for an expanded UK newborn screening programme. This presents a number of new challenges for policy advisors.

Multiplex testing has been available for years. Tandem mass spectrometry, currently used in the newborn blood spot (NBS) screening programme, can detect dozens of statistical abnormalities in the blood spot. Each condition included on the NBS programme and any potential candidate conditions have been assessed in turn. However, there is pressure to assess all the conditions that might be found with WGS at once.

The genotype-first approach has the potential for harms, and in some cases, may be more uncertain than more traditional methods. Not everyone with a pathogenic variant will develop symptomatic disease (incomplete penetrance), and symptoms caused by the same genetic variant can vary in severity among affected people (expressivity).

A cost-effectiveness analysis of WGS in newborn screening will be needed for a policy decision, but screening for potentially hundreds of conditions with a single test will require a different methodological approach than one that focuses on a single condition.

Finally, the use of WGS for newborn screening presents several ethical challenges. The majority are common to all screening programmes (anxiety, informed choice and penetrance), but there are some that are more pressing or likely in this programme. For example, some of the genetic variants might only be of significance later in life, and there are implications for the relatives if a variant is found and there is considerable commercial interest in secondary uses of the data which will not benefit participants directly.

We, therefore, aimed to (1) assess different evidence sources and approaches to evidence synthesis, (2) review methods for evaluating cost-effectiveness and (3) collate views of the public on the main challenges of WGS to inform an approach to assessing WGS for newborn screening in the future.

Objectives

1. To undertake a series of five systematic reviews covering a stratified (by burden and cost of the intervention) random sample of rare diseases to establish the evidence base per condition and to provide a reference case for comparison with alternative review approaches. The reviews addressed six questions mapped to the UK National Screening Committee (NSC) criteria on penetrance and expressivity, the proportion of children with disease who carry gene variants, test accuracy, effectiveness of earlier treatment, effectiveness and benefits and harms of WGS.
2. To explore the utility of the existing online resource Clinical Genome Resource (ClinGen) to provide evidence on the actionability of rare paediatric genetic diseases in order to evaluate it as a potential evidence source for the UK NSC.
3. To undertake a review of genomic studies of newborn screening cohorts reporting penetrance of pathogenic variants to explore the feasibility of identifying highly penetrant pathogenic variants that could be considered for a screening programme.
4. To produce a methodological overview of existing published economic evaluations and costing studies of WGS or whole exome sequencing (WES).
5. To explore patient and public views about the introduction of WGS for newborn screening.

Methods

Review of five conditions

A stratified random sample of five conditions was reviewed. Stratification was based on a range of scenarios that might reasonably have an impact on the UK NSC's recommendations relating to WGS for newborn screening. The five conditions were:

a. pyridoxine-dependent epilepsy (PDE)
b. heritable retinoblastoma (hRB)
c. X-linked hypophosphataemic rickets (XLHR)
d. familial haemophagocytic lymphohistiocytosis
e. medium-chain acyl-CoA dehydrogenase deficiency (MCADD).

Data sources

MEDLINE (via Ovid), EMBASE (via Ovid), Science Citation Index (SCI) (via Clarivate) and the Cochrane Library (via Wiley) from inception to November 2023.

Study eligibility criteria were defined for each review question and included the following:

Population

This included the below:studies of newborn screening cohorts, or
studies of newborns and children (≤ 18 years) with clinical, or biochemical and clinical features of the five conditions.

Exposure/intervention

This included the below:presence of pathogenic variants in the relevant gene(s) detected by sequencing
eligible interventions relevant to the screening context, with 'early' intervention defined separately for each condition.

Target condition

This included the below:clinically or clinically and biochemically defined disease.

Outcomes

These included the below:
measures of disease-specific morbidity and mortality
any health-related health outcomes that could be measured across conditions
any harms or other benefits from WGS.

We produced a narrative synthesis of studies.

Exploring Clinical Genome Resource as an evidence source

We searched the ClinGen database on 19 February 2024 for each of the genes included in the review of five conditions, and we tabulated actionability scores and evidence levels comparatively against the evaluation from GEL and against our assessment using the UK NSC criteria.

Review of genomic studies of newborn screening cohorts reporting penetrance of pathogenic variants

Data source: MEDLINE (via Ovid), EMBASE (via Ovid), SCI (via Clarivate) and the Cochrane Library (via Wiley) from inception to January 2024.

Study eligibility criteria: Studies of unselected newborns sequenced for any rare condition with outcomes of penetrance or an approximation.

We produced a narrative synthesis of our findings.

Methods for review of cost-effectiveness evaluations of whole genome sequencing and whole exome sequencing

Data source: MEDLINE (Ovid), EMBASE (Ovid), cost-effectiveness analysis registry, Web of Science (WoS) and American Economic Association electronic bibliography (EconLit) from inception to February 2024 and hand searches of identified systematic reviews.

Study eligibility criteria: Economic evaluations, clinical trials and Health Technology Assessments reporting costs of WGS or WES in human health care.

Evidence synthesis: a general narrative synthesis of the methodological approaches adopted will be reported. We will also focus on two specific methodological questions:

1. How were the costs associated with WGS and WES estimated?
2. What comparators were included in each study?

Consideration of the public voice in the evaluation of whole genome sequencing

Eight members of the public attended five 2-hour virtual meetings between 15 January and 21 May 2024. Meetings were deliberative and pre-defined topics related to WGS were explored: harms and benefits, genetic uncertainties, systematic review findings and the role of patient and public involvement and engagement (PPIE) in future reviews.

Themes of participants' views were narratively synthesised.

Results

Review of five conditions

Extrapolating the traditional approach to 200 conditions

We screened 19,689 titles and abstracts for the 5 traditional reviews, of which 1348 were selected for full-text assessment (range 55–449 per condition). A total of 268 studies were eligible for inclusion across the 5 reviews (range 31–78). No evidence was identified for the four review questions that required studies to be conducted in newborns. Overall, the five traditional reviews yielded very little of the evidence required by the UK NSC. Considering the time taken to identify and select the evidence, and extrapolating to a review of 200 conditions, we could expect as many as 787,560 unique records, 53,920 full texts to be screened and 8840 studies to be reviewed and synthesised, which is estimated to take a team of 5 reviewers 23 years to complete.

Evidence on the genetic spectrum in children with disease

Two hundred and sixty studies (range 26–73) were included, which reported the genetic spectrum in children with the five conditions. The proportion of children testing positive on sequencing varied for each condition by:

1. definition of disease from broadest (symptomatically defined) to narrowest (genetically defined) category
2. the testing strategy (type of test, number of genes and extent of gene sequencing and additional genetic testing to supplement sequencing)
3. the extent of 'pre-screening' using biochemical and clinical markers.

At variant level, studies provided data on the proportion of novel variants and type of variants but very little information on the severity of disease for specific variants. The large number of novel variants present a challenge to sequencing newborns as their pathogenicity is difficult to ascertain.

Evidence on early versus late treatment

Twenty-two studies (range 1–9) reported the outcomes of early versus late treatment. No study was designed to compare treatment effectiveness in screen-detected versus symptomatically detected children. Definitions of early and late treatment varied and relied on study authors' definitions. The evidence base pointed towards some benefit in

early treatment. However, the quality and volume of the evidence were low because of the definition of early versus late, the type of study, the number of participants and the number of studies available. Therefore, there was insufficient evidence to clearly judge the effect.

Learnings from the five traditional reviews

A single approach to reviewing five conditions was not feasible due to differences in the conditions' characteristics, treatment and aim of screening. For instance, each search strategy was developed individually. Disease-specific categories were needed to organise studies by the population subtype because of differences in the availability of biochemical tests, the number of disease groups with overlapping symptoms and whether conditions could only be defined genetically. The definition of early versus late depended on whether the relevant intervention was preventative, curative or for symptom management, whether an early intervention phase could be defined and whether conditions were progressive or presented following a trigger. A review of 200 conditions would require 200 individual reviews; however, some learning may be transferable between reviews of similar conditions, which we could not explore with the 5 conditions.

Exploring Clinical Genome Resource as an evidence source for the United Kingdom National Screening Committee

Four of the five conditions reviewed (PDE, hRB, XLHR and MCADD) had a paediatric actionability report available on ClinGen in March 2024. However, no information on variant classification in terms of pathogenicity was available for any of the genes.

Comparison of our assessment of the five conditions using the UK NSC criteria with the ClinGen scores of actionability alongside GEL's decisions to include genes on their gene list was complicated. The overall decision on actionability differed for four/four conditions between ClinGen and our assessment using the UK NSC criteria and for five/five conditions between GEL's assessment and our assessment.

It would be inappropriate for the UK NSC to base decisions on potential screening programs on the actionability reported in ClinGen without further assessment.

Review of genomic studies of newborn screening cohorts reporting penetrance of pathogenic variants

Fourteen studies reported experiences with gene sequencing in newborns, of which five provided information that approximated penetrance by reporting some clinical follow-up after a sequence positive test. The number of included genes ranged from 134 to 954 across the five studies and the number of newborns sequenced ranged from 127 to 29,989. Gene selection and variant interpretation varied across studies.

The proportion of babies designated as screen positive from these studies ranged from 1.7% to 9.7%. Half of the positive screens were for conditions not included on conventional newborn screening panels in the study countries (USA and China). However, the clinical significance of a large majority of 'positive screen' results on sequencing (83.3–100%) was unknown, so we do not know if detecting and reporting these was overdiagnosis of clinically insignificant disease, misdiagnosis of disease or early detection of late-onset disease.

Follow-up ranged from 2 months to > 5 years. Penetrance was approximated by the number of confirmed cases after clinical follow-up. For all genes considered together, penetrance ranged from 1.6% after follow-up of 24–48 months to 58.4% after up to 3-year follow-up. The studies did not provide sufficient evidence to understand penetrance for any genetic variant because:

- The number of infants with a specific condition displaying a range of variants was too low.
- Infants with confirmed genetic disease received management that precludes the estimation of penetrance and expressivity for cases without symptomatic confirmation of disease.
- Clinical follow-up was not long enough to include all childhood-onset cases.

Overall, there was little agreement on what genes should be considered in newborn screening, no indication of how to interpret discordant results from NBS programmes and genetic screening and evidence of overdiagnosis. The studies

demonstrated unequivocally that if WGS was to be introduced without further research, it would cause significant problems.

Review of cost-effectiveness evaluations of whole genome sequencing and whole exome sequencing

Eighty-six studies were included in the review. None of them focused on the use of WGS or WES in a screening context. Under half of studies (*n* = 39/86, 45%) were full economic evaluations, of which only 10 were cost–utility studies, that is, studies which estimated the cost per quality-adjusted life-year (QALY). Most evaluations focused only on the costs and outcomes associated with the diagnostic pathway, avoiding the complexity of capturing the impact of a diagnosis on patient management. Two-thirds of the included studies reported a costing perspective; of which one-third [29/86 (36%)] adopted a broad healthcare system perspective, 15 a specific health system perspective, 8 a patient perspective and 5 a societal perspective. Only seven studies (8%) adopted a lifetime horizon. Of the studies that included a comparator (78/86, 91%), 44 (56%) explicitly stated that the comparator was the current standard of care testing, consisting of a broad range of tests. Different assumptions were made in terms of which tests would no longer be needed following the incorporation of WES or WGS in the diagnostic pathway.

Consideration of the public voice in the evaluation of whole genome sequencing

The group largely supported WGS for newborn screening. As meetings progressed and the complexities were explored, however, views became more nuanced; for example, one participant mentioned that they now were 'sitting on the fence a bit'.

Participants identified a wide range of benefits and harms and broadly felt that the benefits outweigh the harms. Key harms they were concerned about ranged from personal (anxiety) to societal (strain on health services). Key benefits included saving lives and avoidance of a diagnostic odyssey.

The process would have benefited from having more time to develop and discuss ideas. For a future review, it would be beneficial to increase the diversity of viewpoints.

Conclusions

A traditional systematic review approach to evaluating WGS of newborns is unfeasible, and we were unable to identify an acceptable alternative way to evaluate WGS for newborn screening in a single mechanism. Cost-effectiveness evidence for WGS has only focused on symptomatic populations to date. Our review highlights the main evidence gaps and informs the direction of future research efforts.

We propose research undertaken in large joined-up collaborations to produce the evidence that is needed for policy advisors before an evaluation of WGS is feasible. This may include a co-ordinated international approach to collecting penetrance data for pathogenic variants with a clear treatment plan. This could be followed by a staged approach of evaluation considering only those of the 200 conditions for screening that have pathogenic variants with very high penetrance.

Study registration

This study is registered as PROSPERO CRD42023475529.

Funding

This award was funded by the National Institute for Health and Care Research (NIHR) Evidence Synthesis Programme (NIHR award ref: NIHR159928) and is published in full in *Health Technology Assessment*; Vol. 29, No. 65. See the NIHR Funding and Awards website for further award information.

Health Technology Assessment

ISSN 2046-4924 (Online)

Impact factor: 4

A list of Journals Library editors can be found on the NIHR Journals Library website

Launched in 1997, *Health Technology Assessment* (HTA) has an impact factor of 4 and is ranked 30th (out of 174 titles) in the 'Health Care Sciences & Services' category of the Clarivate 2022 Journal Citation Reports (Science Edition). It is also indexed by MEDLINE, CINAHL (EBSCO Information Services, Ipswich, MA, USA), EMBASE (Elsevier, Amsterdam, the Netherlands), NCBI Bookshelf, DOAJ, Europe PMC, the Cochrane Library (John Wiley & Sons, Inc., Hoboken, NJ, USA), INAHTA, the British Nursing Index (ProQuest LLC, Ann Arbor, MI, USA), Ulrichsweb™ (ProQuest LLC, Ann Arbor, MI, USA) and the Science Citation Index Expanded™ (Clarivate™, Philadelphia, PA, USA).

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta.

Criteria for inclusion in the *Health Technology Assessment* journal
Manuscripts are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

Health Technology Assessment (HTA) research is undertaken where some evidence already exists to show that a technology can be effective and this needs to be compared to the current standard intervention to see which works best. Research can evaluate any intervention used in the treatment, prevention or diagnosis of disease, provided the study outcomes lead to findings that have the potential to be of direct benefit to NHS patients. Technologies in this context mean any method used to promote health; prevent and treat disease; and improve rehabilitation or long-term care. They are not confined to new drugs and include any intervention used in the treatment, prevention or diagnosis of disease.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

This article

The research reported in this issue of the journal was commissioned and funded by the Evidence Synthesis Programme on behalf of NICE as award number NIHR159928. The protocol was agreed in June 2023. The draft manuscript began editorial review in July 2024 and was accepted for publication in February 2025. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' manuscript and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this article.

This article presents independent research funded by the National Institute for Health and Care Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, the HTA programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, the HTA programme or the Department of Health and Social Care.

This article was published based on current knowledge at the time and date of publication. NIHR is committed to being inclusive and will continually monitor best practice and guidance in relation to terminology and language to ensure that we remain relevant to our stakeholders.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Newgen Digitalworks Pvt Ltd, Chennai, India (www.newgen.co).