bmjmedicine

Check for updates

Challenges in evaluating whole genome sequencing for newborn screening: series of systematic reviews and roadmap for evidence generation for policy advisers

Karoline Freeman [1], Dylan Taylor,[1] Jacqueline Dinnes [2,3], Corinna C A Clark [1], Inès Kander,[1] Katie Scandrett [2], Shivashri Chockalingam,[1] Naila Dracup,[1] Rachel Court,[1] Furqan Butt,[1] Cristina Visintin [4], James R Bonham,[5] David Elliman [6], Graham Shortland,[4] Anne Mackie,[4] Zosia Helena Miedzybrodzka,[7] Sian Mair Morgan [8], Felicity Kate Boardman [1], Yemisi Takwoingi [2,3], Bethany Shinkins [1], Aileen Clarke,[1] Sian Taylor-Phillips [1]

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/bmjmed-2025-001726).

For numbered affiliations see end of article.
Correspondence to: Dr Karoline Freeman; k.freeman@warwick.ac.uk

Cite this as: *BMJMED* 2025;4:e001726. doi:10.1136/bmjmed-2025-001726

Received: 3 June 2025
Accepted: 24 October 2025

ABSTRACT

OBJECTIVE To evaluate systematic review approaches to synthesising evidence for policy advisers who are considering whether to screen newborns for hundreds of rare diseases using whole genome sequencing.

DESIGN Series of systematic reviews and roadmap for evidence generation for policy advisers.

DATA SOURCES Medline, Embase, Science Citation Index, Cochrane Library.

METHODS 200 conditions included in Genomics England's Generation Study were stratified into five groups and one condition randomly selected from each group using criteria designed to maximise variability and availability of data. 30 systematic reviews were undertaken (five conditions, six review questions) about penetrance, detection rate, accuracy, benefit of earlier treatment, and benefits and harms of screening for the five conditions (search from inception to November 2023). Results were synthesised and reviewer time recorded. Genomic studies of newborn screening cohorts that reported penetrance were systematically reviewed using a non-condition specific approach (search inception to January 2024). The conditions and genes selected for reporting by these studies were identified. ClinGen was explored for synthesising evidence. All approaches were assessed by considering review effort and level and quality of evidence.

RESULTS The five conditions selected for systematic review were pyridoxine dependent epilepsy, heritable retinoblastoma, X linked hypophosphataemic rickets, familial haemophagocytic lymphohistiocytosis, and medium chain acyl-CoA dehydrogenase deficiency. 19 689 titles were screened and 268 papers included that addressed two of six research questions (detection rate and treatment benefit). No studies were identified for the remaining four research questions. Total reviewer time for five conditions was seven months. A team of five reviewers would take over 20 years to conduct similar reviews for 200 conditions. 10 published genomic studies of newborn screening cohorts were identified with a total of 76 268 newborns. The number of conditions screened for varied from 74 to 903, with low concordance (two of 1453 genes were included in all 10 studies). Selection of conditions was primarily based on clinical opinion. All studies reported and acted on genetic findings considered clinically significant, preventing collection of penetrance data.

CONCLUSIONS Current evidence synthesis methods are neither feasible nor fruitful to provide policy advisers like the UK National Screening Committee with the evidence needed to understand the benefits and harms of newborn screening for multiple conditions using whole genome sequencing because the evidence is not available to synthesise. Evidence should be created through studies that only report pathogenic variants to parents and clinicians where penetrance and expressivity have been established through empirical evidence rather than clinical opinion. A roadmap for future evidence generation is proposed to highlight the need for new evidence generation combined with a staged approach to evaluation, focusing on pathogenic variants with evidence of high penetrance and expressivity.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Whole genome sequencing is a potentially transformative tool for the detection of rare monogenic diseases
⇒ 15 genomic projects to sequence a total of 400 000 healthy newborns have been started worldwide to evaluate genetic sequencing as a screening tool for hundreds of genetic disorders
⇒ Methods are needed for synthesising evidence on the benefits and harms of using whole genome sequencing to simultaneously screen for hundreds of conditions

WHAT THIS STUDY ADDS

⇒ Using systematic review methods to synthesise the evidence for hundreds of conditions would take a team of five reviewers more than 20 years and would not produce the evidence needed for policy making because the evidence is not available to synthesise
⇒ Alternative approaches are hampered by lack of evidence

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE, OR POLICY

⇒ A roadmap for evidence generation is proposed that could deliver the high quality evidence needed for contemporaneous policy making

REGISTRATION Review registration PROSPERO (CRD42023475529).

Introduction

Whole genome sequencing (WGS) has been suggested as a transformative tool in the detection of rare monogenic childhood diseases. Each rare monogenic disease typically affects less than one in 2000 people and is caused by a single gene variant. Although individually rare, approximately 300 million people worldwide live with rare diseases. Around 80% of rare diseases have a genetic cause and almost 70% of those present in childhood.[1] Although about 95% of rare diseases lack approved treatments,[1] for approximately 600 of about 12 000 childhood onset monogenic diseases there is a potential intervention.[2 3] Early detection may be crucial to improve outcomes. WGS offers potential to identify genetic variants associated with hundreds of childhood onset conditions simultaneously by examining the entire DNA sequence. It could theoretically allow an expansion of current newborn screening programmes as a first line screening test to identify these rare diseases before symptoms develop. WGS also presents unique challenges because of an additional need to understand the presumed association of genetic variants with disease. This involves a shift in concept from current newborn screening that detects biomarkers that are themselves manifestations of the disease.

Interpreting the clinical significance of genetic variants for future health is fraught with uncertainty. Every person has millions of genetic variants when compared with the reference genome.[3] Although some variants are pathogenic, most are benign or of uncertain significance. A meticulous process of systematic evaluation and classification of genetic variants is required to distinguish between disease causing variants and harmless variants as part of gene variant curation efforts for population screening. This process includes assessing the clinical significance of each variant through comprehensive analysis of scientific literature, population frequency data, functional studies, and computational predictions and collaboration among geneticists, bioinformaticians, and clinicians. The goal is to differentiate between pathogenic (or likely pathogenic) variants that may warrant clinical intervention (defined as causing disease in at least 90% of people)[4] and benign variants that do not affect health.

The current classification of variants is primarily based on studies of people who are clinically affected or those with a family history who are more likely to carry the most pathogenic and penetrant variants. This classification needs to be reassessed for the screening context because this will have implications for benefits and harms of WGS for newborn screening (frequently termed genetic newborn screening—gNBS). Predicting disease presentation accurately is difficult, even for pathogenic variants with a well characterised genetic cause, because of incomplete penetrance (not all people with the same pathogenic variant show symptoms of the disease) and variable expressivity (the degree to which a genotype is expressed in a person's phenotype). Reporting pathogenic variants that have low penetrance, reduced expressivity, or are associated with adult onset disease will result in a situation in which many children identified as at risk might undergo medical surveillance, treatments, or lifestyle changes without benefits.[5] Additional variants identified through screening with no previous evidence on pathogenicity (variants of uncertain significance), which should not be disclosed to parents of newborns, could result in children with symptoms not being identified.

In the UK, the Generation Study, using WGS to sequence 100 000 newborns to screen for over 200 rare diseases (covering 478 genes), is under way to examine the potential for an expanded UKNBS programme.[6] New screening programmes, however, require a robust, evidence based understanding of the benefits and harms.[7] Currently, we are not aware of any comprehensive reviews rigorously evaluating the balance of benefits and harms of implementing gNBS, and methods are needed to do this for hundreds of conditions at once. The traditional approach for evaluating evidence for a new screening programme is to conduct a review of published evidence and a cost effectiveness analysis of the programme. Established processes, like those used by the UK National Screening Committee (UKNSC), have been developed for single conditions to assess the volume, direction, and quality of evidence for a number of criteria that allow recommendations to be formulated. In the UK, this involves determining whether the 20 UKNSC criteria relating to the condition, test accuracy, the advantages of early treatment, the harms and benefits of screening, and the programme's ethics, acceptability, feasibility, and cost effectiveness have been met or not. This traditional approach is unlikely to be feasible for hundreds of conditions.

Therefore, we compared the standard UKNSC approach, applied to five of the conditions included in the Generation Study,[8] with an alternative approach that is not condition specific and an online resource of curated evidence on genetic conditions. In our first approach, we sought to establish the evidence base for each of the five conditions and to gauge the feasibility of extending this traditional approach to hundreds of conditions.[9] The second approach to evaluating the evidence for hundreds of proposed conditions aimed to focus evaluation efforts on conditions that have pathogenic variants with evidence of high penetrance (similar to a high positive predictive value of a genetic risk factor). Any

lack of evidence on penetrance for pathogenic variants associated with a condition under consideration could be used to rule out conditions from evidence review. This is because penetrance relates to UKNSC criterion number 1: "The development from latent to declared disease should be understood and/or there should be robust evidence about the association between the risk or disease marker and serious or treatable disease."[10] Ruling out conditions from evidence review based on penetrance evidence may allow a stepped approach to evaluation in the future.

In our third approach, we evaluated the US National Institute of Health funded Clinical Genome Resource (ClinGen). ClinGen is an open access and centralised resource of clinically relevant genes, and ClinGen's paediatric working group uses a standardised protocol to produce summary actionability reports and semi-quantitative metric scores of actionability for childhood onset rare genetic disorders.[11] Clinical actionability is defined as knowledge of penetrance combined with the extent to which available interventions can be used to mitigate the effect of the disease.[12] Although ClinGen aims to define the clinical relevance and actionability of secondary findings in paediatric patients undergoing clinically indicated diagnostic testing, their paediatric protocol states that elements relevant for population based screening decisions may be captured[11] without consideration of systems based practice and availability of population scale follow-up. We assessed whether the evidence summaries and actionability scores in ClinGen's paediatric actionability reports may provide an evidence source for a future review of genomic screening of 200 conditions by comparing them to our assessment of the evidence using the UKNSC criteria.

The aim of this research was to assess these three approaches to synthesising evidence for the evaluation of gNBS and to scrutinise available evidence from different sources for informing policy decisions. The full report is available elsewhere.[9] In this paper, we provide an overview of the methods, findings, and limitations of each approach when applied to gNBS, and discuss the challenges for decision makers.

This research was undertaken in collaboration with and for the UKNSC to inform a balanced and evidence based response to the expected need to evaluate gNBS by defining the evidence that policy makers will need to make decisions about what should be included in a gNBS programme in the future. Our approach was from the public health perspective, focused on understanding the evidence about the balance of benefit and harm for the population offered screening. This approach is of international importance. Several international initiatives have begun exploring the feasibility of genetic screening for newborns,[13] and many national and private providers will have to consider what gNBS could offer for screening.

Methods

We followed a carefully designed protocol prespecifying three approaches to evidence synthesis of benefits and harms of gNBS (PROSPERO CRD42023475529). Firstly, a condition focused set of systematic reviews of five conditions; secondly, a non-condition specific review of genetic newborn screening studies reporting penetrance; and thirdly, an evaluation of ClinGen, an existing resource providing evaluations of actionability of genetic childhood diseases. We assessed the approaches by considering their feasibility for evaluation of more than 200 conditions, and the level and quality of evidence they deliver. Fundamental to the evaluation of gNBS and to the design and delivery of our research was patient and public involvement and engagement (PPIE). We briefly describe our methods for the three approaches and for PPIE below.

Review of five conditions using a traditional approach to evaluation

We randomly sampled five conditions for systematic review. Random selection was from conditions considered to meet the four principles of Genomics England's Generation Study (GEL score 1, judgment in July 2023)[8] and to have high prevalence within the context of rare diseases. Random selection was stratified based on a range of five scenarios (online supplemental material 1) that might reasonably have an impact on the UKNSC's recommendations relating to gNBS. These scenarios were primarily driven by the nature of the interventions that could be introduced on detection of the condition (risk of harm, cost and impact in the short and long term).[9] The five randomly chosen conditions were pyridoxine dependent epilepsy, heritable retinoblastoma, X linked hypophosphataemic rickets, familial haemophagocytic lymphohistiocytosis, and medium chain acyl-CoA dehydrogenase deficiency.

Data sources for the review of the five conditions included Medline (through Ovid), Embase (through Ovid), Science Citation Index (through Clarivate), and the Cochrane Library (through Wiley) from inception to November 2023. We developed five broad search strategies.[9] Study eligibility criteria were defined for six review questions (table 1 and online supplemental material 2 for detailed review questions using the PI/ECOS (population, intervention/exposure, comparator, outcome study) framework) that deal with key UKNSC criteria on penetrance, detection rate, test accuracy, benefits of earlier treatment, benefits and harms, and effectiveness of WGS. We included studies of newborn screening cohorts or of newborns and children (age ≤18 years) with clinical, or biochemical and clinical features of the five conditions. Studies were eligible if the presence of pathogenic variants in the relevant genes was detected by sequencing. For the question of benefits of early treatment, we defined "early" separately for each

Table 1 | Overview of studies screened and included for evidence reviews of five conditions

Screened and included studies	No of studies
Titles and abstracts screened	19 689
Full texts screened	1357
Included studies	
Total	268
Question 1: Penetrance*	0
Question 2: Sensitivity of WGS in children with disease†	245
Question 3: Test accuracy‡	0
Question 4: Early *v* late treatment§	23
Question 5: Effectiveness of WGS¶	0
Question 6: Harms of WGS**	0
Approximate time taken for evidence review	7 months

Data are number of studies unless stated otherwise.

*Question 1. What is the penetrance and expressivity of different gene variants associated with each condition? (Related UKNSC criteria 1 and 3.)

†Question 2. What is the prevalence of genetic variants in those with biochemical or biochemical and clinical features of each condition? (Related UKNSC criterion 1.)

‡Question 3. What is the diagnostic accuracy (clinical validity) of gene sequencing for each condition? (Related UKNSC criteria 4, 5, and 8.)

§Question 4. What is the effectiveness of earlier versus later intervention (treatment or surveillance) for each condition? (Related UKNSC criteria 9 and 10.)

¶Question 5. What is the effectiveness of WGS for newborn screening for each condition? (Related UKNSC criterion 11.)

**Question 6. What are the harms of WGS for newborn screening for each condition and any additional benefits beyond those from earlier treatment? (Related UKNSC criterion 13.)

UKNSC, UK National Screening Committee; WGS, whole genome sequencing.

condition by considering the aim of screening for each condition (prevention, surveillance, early treatment). We produced a narrative synthesis of included studies.

Review of genomic studies in newborn screening cohorts reporting penetrance

We undertook a review of genomic studies of newborn screening cohorts reporting penetrance of pathogenic variants. We searched Medline (through Ovid), Embase (through Ovid), Science Citation Index (through Clarivate), and the Cochrane Library (through Wiley) from inception to January 2024. Search strategies are available in a previous publication.[9] We included studies of unselected newborns sequenced for any rare condition with outcomes of penetrance or an approximation.[9] We produced a narrative synthesis of findings on the proportion of people with positive genetic screening results identified to have confirmed disease on clinical follow-up. We considered positive genetic screening results as unconfirmed if, within the type and length of follow-up (confirmatory testing, clinical follow-up, or both), the condition could not be confirmed.

We also assessed the extent of overlap of gene lists from published newborn genomic studies to gauge consensus on which genes to report to parents of newborns. We used the R (version 4.3.1)[14] package eulerr (version 7.0.2)[15] to draw a Euler diagram to illustrate the overlap of genes considered for reporting across the different studies and the UpSetR package (version 1.4.0)[16] to create an UpSet plot to simultaneously show the number of genes and the number of unique genes for each study.

Evaluation of ClinGen to provide evidence base for policy advisers

We searched the ClinGen database on 30 April 2024 for actionability reports for each of the genes included in the review of five conditions, and actionability scores for all 478 genes on the gene list from the Generation Study (April 2024).[17] Actionability reports summarise the evidence of genetic conditions under four dimensions: severity of disease, likelihood of disease (similar to penetrance), effectiveness of intervention, and nature of intervention. The dimensions are scored from 0 to 3 (3 being best for actionability) based on the evidence reviewed, and the evidence is rated from poor evidence to substantial evidence. The scores are summarised across the four dimensions to provide an overall score. The scores and the evidence are taken into consideration for the final assertion on actionability. We mapped the dimensions to relevant UKNSC criteria to compare the individual and final scores with our assessment of the evidence for the five conditions against the UKNSC criteria (online supplemental material 3). We produced a narrative of the scope and level of evidence informing the actionability scores of the ClinGen resource in comparison to our assessment using the UKNSC criteria and the current gene list of the Generation Study.

Patient and public involvement

We recruited eight people to a PPIE group, six of whom were directly affected by rare diseases (five were parents of children with a rare disease and one was an adult living with a rare disease), who attended five, two hour virtual meetings between 15 January and 21 May 2024 organised by a dedicated PPIE researcher. Meetings were deliberative and explored predefined topics related to evaluation of gNBS: harms and benefits, genetic uncertainties, systematic review findings, and the role of PPIE in future reviews. Important themes from the PPIE work are reported in each results section and are documented in detail in online supplemental material 4.

Results

Lack of evidence from traditional review approach by condition

For our first approach, we screened 1357 full texts across five conditions but found no studies addressing four of the six pivotal questions we investigated to assess nine UKNSC criteria required for policy decisions (table 1). There was no evidence on penetrance of pathogenic variants identified through

screening; no evidence on the test accuracy of WGS in the newborn screening setting; and no studies on the clinical effectiveness, harms, and additional benefits associated with WGS in newborns. Therefore, using the traditional condition by condition approach, we found no studies that investigated gene sequencing in unselected newborns without family history or symptoms.

We included 23 studies (ranging from one to nine across the five conditions) on early versus late treatment. However, in all studies, "early" approximated screen detected disease; none of the studies was designed to compare treatment effectiveness in screen detected versus symptomatically detected disease. Therefore, we do not know whether treatment of newborns with a genetic risk factor is better than treating symptomatic disease in infants and how many might be overtreated. The evidence identified is less robust than that usually required to meet the UKNSC criteria for which there should be an effective intervention for patients identified through screening, with evidence that intervention at a presymptomatic phase leads to better outcomes for the screened person compared with usual care.

We included 245 studies (26-73 studies across the five conditions) reporting sequencing in children with confirmed disease to estimate the proportion of patients missed with WGS.[9] Despite the large volume of evidence, we could not determine the detection rate of WGS from these studies.[9] Firstly, sequencing was frequently supplemented by additional testing to increase the sensitivity of testing. Secondly, four of the five conditions reviewed were defined by symptoms common to other conditions where sequencing resulted in a host of alternative genetic diseases (lack of consistent case definitions).[18] Thirdly, sequencing results in children with abnormal biomarkers (biochemically defined disease) differed from sequencing results in clinical cohorts (symptomatically defined disease). For example, in medium chain acyl-CoA dehydrogenase deficiency the pathogenic c.985A>G variant frequently reported in children with symptomatic disease was found to be less frequent in babies with molecular risk factors in whom "milder" variants were detected and had reduced penetrance than estimated from clinical cohorts.[19] Knowledge on the genetic spectrum obtained from clinical cohorts does not necessarily apply to the screening context, a finding that is likely to be exacerbated when sequencing unselected newborns. Finally, the studies showed that WGS may work better in some conditions than in others. We observed large numbers of new variants (65% of variants detected)[20] to 77%[21] which, for two conditions (pyridoxine dependent epilepsy and X linked hypophosphataemic rickets), occurred only in a small number of people. Considering the known challenges in assessing pathogenicity and penetrance, the potential for large numbers of new variants of unknown significance resulting from screening would negatively affect the sensitivity of WGS and create considerable uncertainty. The possibility of being left with uncertainty rather than clarity after screening and the fact that not all disease will be identified may disagree with public expectation and was a major concern in our PPIE group. This concern developed over time with growing appreciation of the complexities of WGS (online supplemental material 4).

Extrapolating the person days required to conduct these reviews of five conditions to a scenario in which WGS is used to detect 200[8] or even 600 conditions[22] would take a team of five reviewers about 23 and 69 years, respectively. For 200 conditions, the total number of records to screen would be about 787 560 titles and abstracts, 54 280 full texts, with synthesis of 10 720 studies, and for 600 conditions, 2 363 680 titles and abstracts, 161 760 full texts, and 26 520 studies for synthesis.

Overall, using condition based systematic reviews to examine the evidence for using WGS to screen for rare conditions was neither fruitful nor practical. The impracticality of systematic reviews in this context underscores the need for innovative approaches to efficiently identify and synthesise evidence, ensuring timely and informed policy making.

Lack of penetrance estimates from published genomic studies of newborn cohorts for non-condition specific approach to evaluation

For our second approach, we screened 4970 records and identified 14 studies (16 references)[23–38] that reported experiences with newborn gene sequencing. Of these studies, four[35–38] only investigated one condition that is not the use case under consideration and the studies are therefore not discussed further. The remaining 10 studies included a total of 76 268 newborns and between 134[30] and 954[23] genes for reporting (74-903 conditions).[23 30] The studies reported diagnostic yield (the number of newborns with positive screening results divided by the total number of newborns screened) with insufficient follow-up to identify health outcomes to inform an assessment of benefits and harms of WGS in newborn screening. Therefore, the studies were insufficient to assess our staged approach to evaluation, where full evaluation would be based on the availability of evidence on penetrance.

We approximated penetrance for five studies that specified short follow-up of newborns with positive screening results.[23 26 28–30] The proportion of infants with confirmed disease for all genes combined ranged from 1.6% after 24-48 months of follow-up[26] to 58.4% after up to three years of follow-up (online supplemental material 5).[30] The evidence was insufficient to reliably estimate penetrance for any specific pathogenic variant owing to several factors:

- The number of infants with a specific and named variant was generally too low across studies and conditions.
- Infants with confirmed genetic disease received management that precluded estimation of penetrance and expressivity for those without symptomatic confirmation of disease
- Clinical follow-up was not long enough to include all childhood onset diseases.

However, even with the relatively short follow-up, the studies produced evidence of overdiagnosis (online supplemental material 5). Overall, more than 50% (1449/2687) of those identified could not be confirmed by the studies' follow-up methods—often because of lack of follow-up to clinical disease, but also owing to incorrect variant interpretation leading to overdiagnosis.

Comparison to conventional biochemical or symptomatic screening without follow-up (reported in four of 10 studies) did not compensate for short follow-up in most cases because agreement between conventional screening and genetic screening was generally poor, without any indication of how to interpret discrepancies. Furthermore, many conditions predicted were not included in the conventional screening panels (online supplemental material 5).

Lack of agreement on clinical significance of genes for reporting for any approach to evaluation

Concordance in gene selection between the 10 studies was low (figure 1). Only two of 1453 genes were included in all 10 studies (*BTD* associated with biotinidase deficiency and *IL2RG* associated with severe combined immunodeficiency). Half (766 of 1453; range 0-386) of all genes were included as clinically significant in only one of 10 studies (figure 2) and only two studies with small gene lists were devoid of unique genes.[27 30] The processes for gene selection varied greatly across studies.[9] This variation cannot be fully explained by different populations, healthcare perspectives, and motivations for testing, but indicates substantial disagreement in the genes deemed actionable across these studies and uncertainty about what should be reported to parents of newborn babies. Furthermore, the studies were not large enough to produce evidence for individual rare conditions (eg, X linked hypophosphataemic rickets incidence one in 20 000-70 000).[39] The use of different sequencing methods further signifies a lack of consensus on which testing method to use in newborn screening programmes. Understanding these discrepancies requires a thorough examination of gene or condition selection processes, variant annotation and interpretation, and calling rules used by different

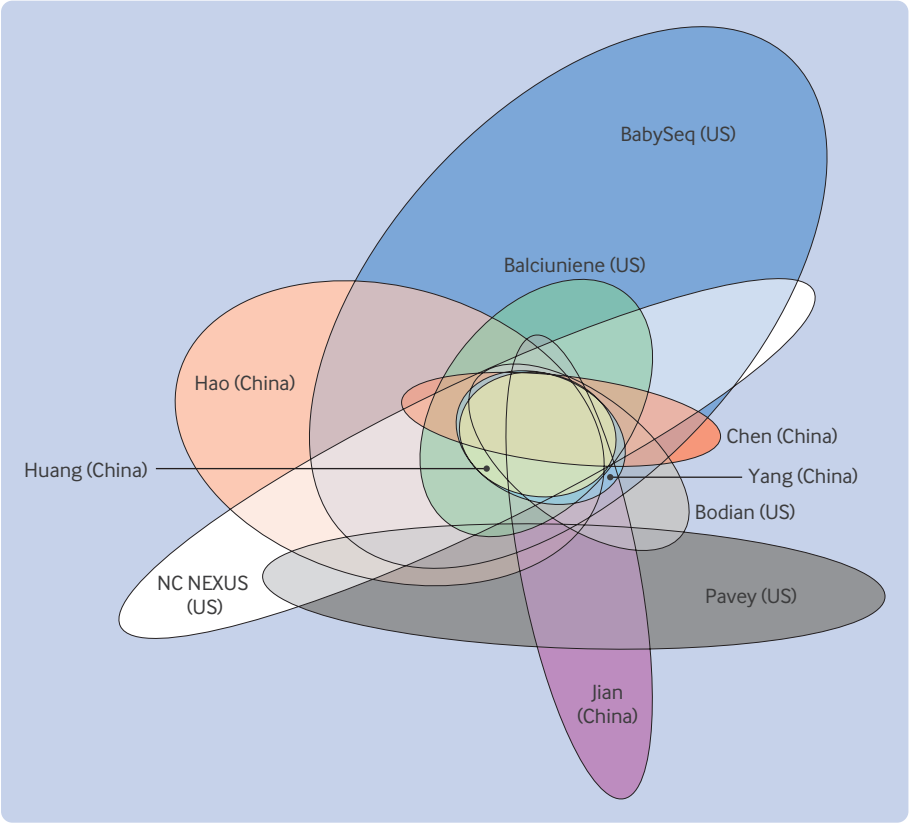

Figure 1 | Euler diagram showing lack of overlap of genes included in 10 published studies evaluating genetic screening for newborn babies

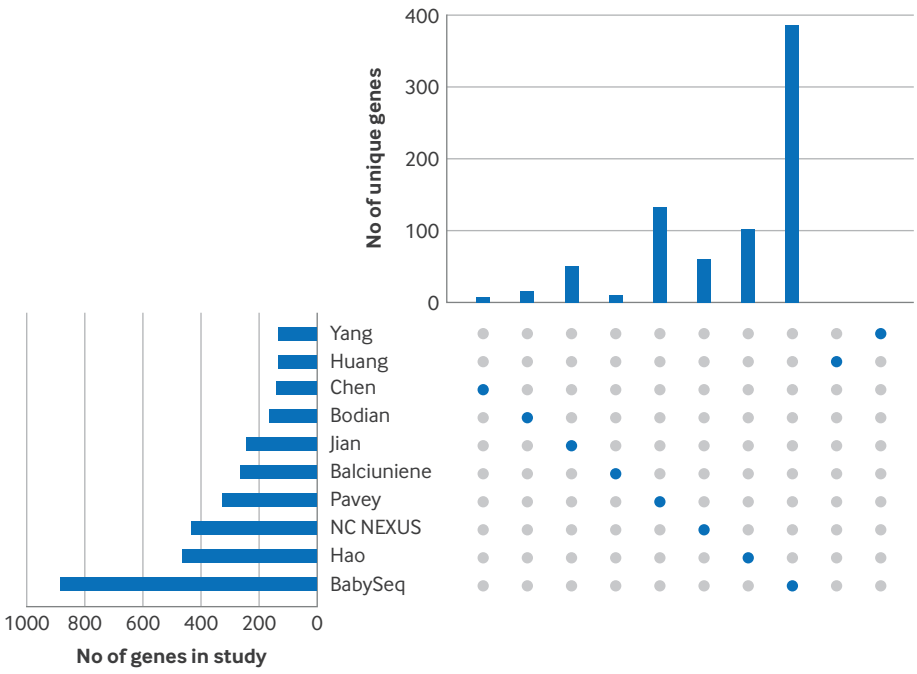

Figure 2 | UpSet plot showing number of unique genes by study (evaluating genetic screening for newborn babies) and their respective total number of genes selected for reporting

studies and a better understanding of differences in clinical opinions.

Overall, the 10 studies highlight that gNBS is still in its infancy and precludes a conclusive evaluation of the feasibility of the approach to gNBS evaluation that focuses on penetrance. There was a lack of consensus on which genes to consider for newborn screening, no indication of how to interpret discordant results from conventional newborn screening programmes and genetic screening, uncertainty over which test is most appropriate for newborn screening, and a large number of positive results that could not be confirmed within the studies' follow-up, resulting in interventions, routine surveillance, and regular follow-up of uncertain benefit. Our PPIE group raised concerns over unnecessary worry and parental anxiety that would be created with unconfirmed positive screening results, and over the quality of education for parents about what screening results mean for their families (online supplemental material 4).

Insufficient evidence from international gene curation efforts to use existing evidence synthesis products for decision making

Our third approach evaluated the ClinGen resource. Four of the five conditions reviewed (pyridoxine dependent epilepsy, heritable retinoblastoma, X linked hypophosphataemic rickets, and medium chain acyl-CoA dehydrogenase deficiency) had a paediatric actionability report available on ClinGen in April 2024. However, ClinGen provided no information on pathogenicity and penetrance at variant level for any of the disease causing genes. The overall decision on actionability differed for four of four conditions between ClinGen and our assessment from the five systematic reviews using the UKNSC criteria.[9] Taking into account the severity of the condition, penetrance, and treatment effectiveness in children without symptoms, none of the four conditions met the UKNSC criteria but were rated as highly actionable by ClinGen (online supplemental material 3).

ClinGen covered only 15.5% (74/478) of genes included in the Generation Study (online supplemental material 6, table S1), indicating limited scope as a potential evidence source for evaluating contemporary genomic projects. The summary actionability scores for the 74 genes covered ranged from 3 (low actionability) to 12 (strong actionability) (mean=9.6, median=10). ClinGen genes not included in the Generation Study showed comparable actionability scores (mean=9.7, median=10). Two genes included in the Generation Study did not pass ClinGen's stage one rule-out (ie, not taken forward to full review) owing to insufficient evidence (online supplemental material 6, table S2). Overall, there was considerable lack of agreement on the importance of the genes included for reporting in the Generation Study and included for assessment by ClinGen.

Currently, it would be inappropriate for the UKNSC to base decisions on potential screening programmes using ClinGen actionability reports without further assessment because they were not designed for population based testing. ClinGen's rule out decisions may, however, be useful to rule out conditions for further assessment. This strategy needs further assessment by investigating the reasons for

BOX 1 | ROADMAP FOR EVIDENCE GENERATION

The problem (what we have got)

Knowledge on penetrance and expressivity of pathogenic variants is key to decision making but evidence is lacking.

Current knowledge about rare diseases is generally based on clinical cohorts, family studies, or case control studies with limited applicability to the screening context. The presence of a variant believed to be pathogenic in a person with a family history of symptomatic disease can be highly predictive of phenotypic disease. The same variant detected by population screening in a person with no other risk factors can be poorly predictive of phenotypic disease.

Evidence from screening using biochemical markers suggests that screening cohorts show a wider genetic spectrum, variants associated with milder disease, and greater variation in expressivity.

The genetic spectrum with information on penetrance and expressivity for apparently healthy newborns with a genetic risk factor is unknown.

Current studies are not expected to provide the evidence needed for decision making, because they are too small and report genetic findings to parents with the aim of managing presymptomatic disease. Reporting of genetic findings precludes evidence collection of penetrance, particularly in those with available preventative treatment where symptom development is prevented, and true disease defined by clinical symptoms cannot be confirmed or refuted.

Ways to estimate penetrance and expressivity (what we need)

To estimate penetrance data, the ideal follow-up for each condition needs to be established. For some conditions, where a link between a biochemical marker and clinical disease is well established, a confirmatory biochemical test may be sufficient. In others, follow-up to symptomatic disease needs to be long enough to capture all childhood onset diseases.

Close collaboration with policy advisers to design studies that deliver the evidence that is needed for decision making is essential. Proposed study designs include the following:

⇒ Ideal studies would be large cohort studies where either screening with whole genome sequencing (WGS) is given to newborn babies, without reporting results to parents, or very few conditions of well evidenced penetrance and expressivity are reported. Other potential conditions or genes should not be reported to participants unless they present symptomatically. These studies will allow measuring the penetrance and clinical significance of pathogenic variants and consequently the benefits and overdiagnosis harms of revealing the test results. However, these studies would raise considerable ethical questions and would only be possible with parents' complete understanding, agreement, and consent.

⇒ Gene frequency information from healthy adult cohorts, such as the UK Biobank and worldwide datasets like gnomAD, could be used to identify low penetrance variants that are not suitable for inclusion on a screening panel. However, these studies are less useful for identifying which variants to include because they will exclude people with the disease who have died before reaching adulthood.

⇒ Whole genome or exome sequencing of stored dried blood spot samples could provide data on penetrance. The challenges of doing these studies at scale are the cost and linking to good phenotype data to determine symptomatic disease status.

A staged approach to WGS evaluation (the future)

Once evidence on penetrance and expressivity has been generated, a stepped approach to evaluation may be possible in the future. A full evaluation of benefits and harms can be restricted to conditions associated with genes with at least one pathogenic variant with sufficient evidence on high penetrance and expressivity. Any future, large scale WGS testing and screening programmes need to be established in ways that support robust evaluation, research, and assessment of clinical effectiveness.

disagreement between the Generation Study and ClinGen.

Discussion

Main findings

Current evidence synthesis methods are neither feasible nor fruitful for the evaluation of gNBS because the evidence is not available to synthesise. The current evidence base on gNBS is insufficient to assess the benefits and harms of a newborn screening programme using WGS. As a consequence, there is uncertainty surrounding which genetic diseases should be reported to parents of newborn babies who are asymptomatic, and this is reflected in the differences in gene lists across studies. The condition based systematic review approach is impractical for hundreds of conditions and did not deliver evidence from the screening setting. However, there

is some indication that available evidence from clinical cohorts is not easily applicable to the screening context, which is unsurprising because it is widely accepted that risk estimates for genetic disease from high risk groups do not translate to the general population, as has been shown for cancer predisposition genes in people with and without family history.[40] A staged approach to evaluation, for example, one starting with available evidence on penetrance, is not yet feasible. This approach requires new large studies that allow conclusions to be drawn on variant level, and could include screening cohort studies with long follow-up and without reporting to parents considering their ethical complexities, studies using stored newborn blood spot samples, and studies using biobanks of healthy adults to rule out low penetrance variants (box 1). ClinGen is an open access gene curation resource funded by the National Institute for Health and has different evidence requirements than required for UK policy decisions. The ClinGen protocol explicitly states that it has a relatively low threshold for the type of evidence accepted, including non-systematic or expert based references.[11] The definition of clinical actionability of secondary findings in paediatric patients undergoing clinically indicated diagnostic testing does not translate well to the screening context. ClinGen is unlikely to be a useful evidence source for UKNSC decision making. Evaluation of gNBS is not feasible without further research.

Findings in light of existing studies

As of June 2024, at least 15 ongoing projects have been aimed at sequencing over 400 000 healthy newborns worldwide. Full evaluation of the ongoing projects is premature, however the consensus between gene lists across these projects has been investigated in two recent studies. Downie and colleagues compared gene lists across six of the ongoing genomics projects. They identified 55 genes that were common to all six panels and a further 59 that were on five of the six panels.[41] The authors suggested that the consensus list of 55 genes could be the starting point for harmonisation of gene lists internationally. In contrast, comparison of gene lists across the same six projects and 13 additional published studies showed lower agreement on which gene variants to report to new parents (17/1750 genes (1%) were included in all lists). Our comparison identified only two genes (2/1453, 0.1%) that were included by 10 published studies, including three studies not included by Minten and colleagues.[42] In combination, these studies show overall low levels of consensus, which differ greatly based on which studies are included.

International efforts are under way to harmonise the evidence and resources to responsibly implement newborn sequencing, with a promise to predict treatable disease in babies and intervene before symptoms develop.[13] Minten and colleagues suggested prioritising genes that are included on more than 80% of gene lists for population-wide implementation.[42] However, a consensus list of genes for implementation may be inappropriate and premature for two key reasons. Firstly, the reason for the lack of consensus is the lack of evidence, and reaching consensus will preclude creation of penetrance evidence for any genes on the agreed list. Secondly, evidence indicates that gene selection is skewed towards western Europeans because non-European people are underrepresented in reference population databases used for variant frequency annotation.[43] A gene list based on consensus rather than new evidence could further exacerbate inequalities and deepen health disparities in populations with a broad mix of ethnicities such as in the UK and the US.

Strengths

We attempted to identify an approach to assessing the benefits and harms of gNBS for hundreds of conditions, taking the perspective of policy advisers, and counteracting the push for implementation based on poor evidence. We consulted widely with policy advisers and clinicians throughout the review process to ensure that we framed the right research questions for policy advisers and explored patient and public views on involvement in future evaluations. Our review applied traditional and new approaches, and has the potential to guide policy makers towards a new approach for producing and synthesising the evidence required to make evidence based policy decisions in this complex area.

Limitations

Our review focused on methods to evaluate the clinical effectiveness of gNBS. The limited evidence base meant that we could not deal with several questions that are fundamental to the evaluation of newborn sequencing. Additionally, questions such as the cost effectiveness of gNBS and health related quality of life data for newborn screen detected diseases[44] require further research before they can be addressed for policy decisions.

Our review did not address the question of whether WGS is the most appropriate test for expanded newborn screening compared with whole exom sequencing, panel testing, sequencing combined with other tests, or biochemical assays. A technology centric approach has been identified as inappropriate for population screening,[45] but is actively pursued by gNBS studies. Many conditions could be tested for more appropriately using biochemical assays or bespoke genetic tests that could be superior to WGS.[28 45] WGS is limited in the detection of large deletions and inversions, and further, the use of shallow sequencing, as is frequently implemented in clinical practice for economic reasons, may worsen these limitations so that study results might not be achieved in clinical practice.

Furthermore, we did not assess the impact of sequencing results on clinician behaviour and subsequent clinical care, nor did we look into anticipated workforce challenges with the implementation of WGS as a new screening programme. Challenges include the number of genetic counsellors needed, training of staff to correctly interpret WGS results, and the availability of evidence based clinical pathways for babies with rare asymptomatic genetic disease. Establishing a newborn screening programme using WGS would incur considerable opportunity costs.

Finally, more work is needed to explore key ethical issues around storing blood samples and genetic information (highlighted by our PPIE group as a concern), including for secondary use[46]; masking findings versus knowing but not reporting findings of certain variants; and acceptability of WGS in the light of overdiagnosis, and variable accuracy in different ethnic groups, with additional ethical and equity issues.[47]

Implications for decision makers

Traditional approaches to evidence synthesis are impractical, and traditional and alternative approaches are hampered by lack of evidence. Currently, no evidence supports large scale implementation of gNBS with simultaneous testing of many conditions. The balance of benefit and harm is unknown, and implementation efforts, unless carefully considered, might preclude the research required to measure the benefits and harms. Furthermore, ongoing gNBS studies are not set up to answer these questions,[46] because children with an identified variant will all be treated, and the natural history of the variant will therefore not be determined. The high level of uncertainty surrounding gene selection and the harms and benefits from gNBS, together with ethical and civil liberty concerns, risk eroding public confidence about newborn screening more generally.

Although much is still unknown about the benefits and harms of screening, a helpful approach might be to consider the accurate measurement of penetrance first, so that review efforts concerning other aspects, such as the benefit of earlier treatment, can be targeted to those conditions that have pathogenic variants with sufficient penetrance. This pragmatic stepwise approach is motivated by the paucity of data and the need to engage with the broader scientific community to conduct the types of studies that will deliver the required penetrance data for pathogenic variants. However, there is a widespread belief that population-wide genomic screening will deliver the knowledge on penetrance.[42] Furthermore, reporting genetic findings to families with the goal of early intervention precludes data collection on disease onset and accurate acceptability, because parents are likely to report positively when their child is treated and does not get symptoms of the condition. As such, current efforts to implement sequencing for newborns are a critical risk to being able to answer these questions.

If future large scale programmes for WGS testing and screening are established, they should be delivered in ways that support robust evaluation, research, and assessment of clinical effectiveness to determine immediate and longer term benefits and harms. Future research will also require time intensive, collaborative PPIE engagement considering a greater diversity of experience to map diverse views to social characteristics, backgrounds, and particular lived experiences to facilitate a nuanced understanding of perceived harms and benefits, and help inform policy recommendations and implementation.

We have produced a roadmap for evidence generation (box 1) to refocus discussion on evidence relating to WGS for newborn screening. Our roadmap lays out the shortcomings of the current evidence base and the type of research that is needed to generate new evidence required to support a robust evaluation of WGS.

Conclusions

Current evidence synthesis methods are neither feasible nor fruitful to provide policy advisers such the UKNSC with the evidence needed to understand the benefits and harms of newborn screening for hundreds of conditions using WGS because the evidence is not currently available to synthesise. International gene curation efforts do not provide a source of existing summaries of the evidence base. The current lack of evidence and uncertainty will produce harms (eg, overdiagnosis and overtreatment) if WGS is implemented prematurely. Evidence should be created through studies that only report pathogenic variants to parents and clinicians where penetrance and expressivity have been established through empirical evidence rather than clinical opinion. We propose that any future large scale programmes for WGS testing and screening are established in ways that support robust evaluation, research, and assessment of clinical effectiveness.

AUTHOR AFFILIATIONS

[1]Warwick Medical School, University of Warwick, Coventry, UK
[2]Department of Applied Health Sciences, University of Birmingham, Birmingham, UK
[3]National Institute for Health and Care Research (NIHR) Birmingham Biomedical Research Centre, University of Birmingham, Birmingham, UK
[4]UK National Screening Committee, London, UK
[5]Sheffield Children's NHS Foundation Trust, Sheffield, UK
[6]Great Ormond Street Hospital for Children, London, UK
[7]Division of Applied Medicine, University of Aberdeen College of Life Sciences and Medicine, Aberdeen, UK
[8]All Wales Medical Genomics Service, Cardiff, UK

Acknowledgements We thank the input of the seven members of the patient and public involvement and engagement group: Kerri Pearce, Matt Howard-Murray, Frances Othen-Wales, Luis Canto E Castro, and Helen Bates; and Helen Jenkinson, Austen Worth, Emma Footit, and Raja Padidela for their specialist clinical advice on the rare conditions.

Contributors KF and STP drafted the manuscript. KF, STP, JD, CV, JRB, DE, GS, AM, FKB, YT, and BS contributed to the study conceptualisation and design. KF, JD, IK, KS, SC, and FB undertook the review. ND and RC designed and ran the searches. DT visualised the data. All authors commented on the draft, read and approved the final version. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted. STP acts as the guarantor. Transparency: The lead author (the guarantor) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Funding This study/project is funded by the National Institute for Health and Care Research (NIHR) Evidence Synthesis Programme (ESG_HTA_NIHR159928). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Competing interests All authors have completed the ICMJE uniform disclosure form at www.icmje.org/disclosure-of-interest/ and declare: KF, JD, BS, CCAC, IK, KS, SC, ND, RC, FKB, FB, YT, and STP received funding from the NIHR Evidence Synthesis Programme (ESG_HTA_NIHR1599280) to undertake this work. DT received funding from the NIHR Pre-doctoral Fellowship Programme (NIHR303446) and holds a PhD studentship funded by Novartis. JD and YT were supported by the NIHR Birmingham Biomedical Research Centre (BRC). CCAC and FB received funding from Novartis Gene Therapies to explore the public acceptability of newborn screening for SMA in the UK and Public Health England for the evaluation of Next Generation Sequencing for CF Newborn Screening. FKB received funding from Genomics England Ltd for a process and impact evaluation, MRC/NIHR for a realist evaluation of remote qualitative data collection in health, Wellcome Trust for the development of future research agendas on the socialethical impact of genomics through interdisciplinary live data-sharing, and Diabetes UK for the ELSA study. BS is a member of the UKNSC. CV is a research lead at the UKNSC. DE is clinical adviser to the UKNSC and chair of the UK NSC Blood Spot Task Group. AM is the director of programmes for the UKNSC. JRB is president of the International Society of Neonatal Screening and member of the UKNSC Fetal Maternal Child Health group. GS is vice chair of the UKNSC. ZHM is co-chair of the Scottish Clinical Genomics Forum and a committee member of the Clinical Genetics Society, the Scottish Strategic Network for genomic medicine, the NHSE rare disease test evaluation working group, and board member of the Royal Colleges Genomics Advisory Board. FKB is a member of the UKNSC Foetal, Maternal and Child Health Reference Group, the Bloodspot Task Group and the Spinal Muscular Atrophy In Service Evaluation Partnership Board. STP is a member of the UKNSC AI task group, and chair of the UKNSC Research Methodology Group. SMM and AC have nothing to declare.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information.

ORCID iDs

Karoline Freeman https://orcid.org/0000-0002-9963-2918
Jacqueline Dinnes https://orcid.org/0000-0003-1343-7335
Corinna C A Clark https://orcid.org/0000-0002-1077-9383
Katie Scandrett https://orcid.org/0000-0001-6111-2805
Cristina Visintin https://orcid.org/0009-0006-4250-2924
David Elliman https://orcid.org/0000-0002-8128-2825
Sian Mair Morgan https://orcid.org/0000-0002-4943-653X
Felicity Kate Boardman https://orcid.org/0000-0002-3268-6276
Yemisi Takwoingi https://orcid.org/0000-0002-5828-9746
Bethany Shinkins https://orcid.org/0000-0001-5350-1018
Sian Taylor-Phillips https://orcid.org/0000-0002-1841-4346

REFERENCES

1 The Lancet Global Health. The landscape for rare diseases in 2024. Lancet Glob Health 2024;12:S2214-109X(24)00056-1. 10.1016/S2214-109X(24)00056-1
2 Genomics Education Programe. NHS England London; Newborn screening: time for a genomic approach?, 2021. Available: https://www.genomicseducation.hee.nhs.uk/blog/newborn-screening-time-for-a-genomic-approach/ [Accessed 14 Aug 2022].
3 Auton A, Abecasis GR, Altshuler DM (Co-Chair), *et al*. A global reference for human genetic variation. Nature New Biol 2015;526:68–74. 10.1038/nature15393
4 Richards S, Aziz N, Bale S, *et al*. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015;17:405–24. 10.1038/gim.2015.30
5 Boardman F, Clark C. "We're kind of like genetic nomads": Parents' experiences of biographical disruption and uncertainty following in/conclusive results from newborn cystic fibrosis screening. Soc Sci Med 2022;301:114972. 10.1016/j.socscimed.2022.114972
6 Genomics England. NHS; Newborn Genomes Programme London, 2021. Available: https://files.genomicsengland.co.uk/documents/Newborns-Vision-Final_SEP_2021-11-02-122418_jjne.pdf [Accessed 16 Mar 2024].
7 Downie L, Halliday J, Lewis S, *et al*. Principles of Genomic Newborn Screening Programs: A Systematic Review. JAMA Netw Open 2021;4:e2114336. 10.1001/jamanetworkopen.2021.14336
8 Genomics England. London Genomics England Limited; How we choose conditions, 2024. Available: https://www.genomicsengland.co.uk/initiatives/newborns/choosing-conditions [Accessed 8 Jun 2023].
9 Freeman K, Dinnes J, Shinkins B, *et al*. Evaluating whole genome sequencing for rare diseases in newborn screening: evidence synthesis from a series of systematic reviews. *Pediatrics* [Preprint] 2024. 10.1101/2024.09.03.24312979
10 UK National Screening Committee. UK National Screening Committee; Criteria for a population screening programme, 2022. Available: https://www.gov.uk/government/publications/evidence-review-criteria-national-screening-programmes/criteria-for-appraising-the-viability-effectiveness-and-appropriateness-of-a-screening-programme [Accessed 7 Aug 2024].
11 ClinGen. National Institutes of Health (NIH), Bethesda, MD; ClinGen Pediatric Actionability Workgroup Protocol: Generation of Summary Reports and Semi-Quantitative Metric Scoring, 2020. Available: https://www.clinicalgenome.org/site/assets/files/5075/peds_combined_evidence_curation_and_scoring_protocol_07292020.pdf [Accessed 7 Mar 2024].
12 Wright CF, West B, Tuke M, *et al*. Assessing the Pathogenicity, Penetrance, and Expressivity of Putative Disease-Causing Variants in a Population Setting. Am J Hum Genet 2019;104:275–86. 10.1016/j.ajhg.2018.12.015
13 ICoNS. We are icons the international consortium on newborn sequencing. In: Consortium on Newborn Sequencing (ICoNS). London, 2024. Available: https://www.iconseq.org/ [accessed 14 May 2024].
14 Team RC. R: a language and environment for statistical computing. r foundation for statistical computing. *Vienna*; 2023. Available: https://www.R-project.org/ [accessed 5 Feb 2024]
15 Larsson J. eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses. R package version 7.0.2, 2024. Available: https://CRAN.R-project.org/package=eulerr [Accessed 15 Jan 2024].
16 Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics 2017;33:2938–40. 10.1093/bioinformatics/btx364
17 Genomics England. Conditions list London: Department of Health and Social Care, 2024. Available: https://www.genomicsengland.co.uk/initiatives/newborns/choosing-conditions/conditions-list-generation-study [Accessed 9 Mar 2024].
18 Bick D, Ahmed A, Deen D, *et al*. Newborn Screening by Genomic Sequencing: Opportunities and Challenges. Int J Neonatal Screen 2022;8:40. 10.3390/ijns8030040
19 Andresen BS, Lund AM, Hougaard DM, *et al*. MCAD deficiency in Denmark. Mol Genet Metab 2012;106:175–88. 10.1016/j.ymgme.2012.03.018

20 Jiao X, Xue J, Gong P, *et al*. Clinical and genetic features in pyridoxine-dependent epilepsy: a Chinese cohort study. Dev Med Child Neurol 2020;62:315–21. 10.1111/dmcn.14385
21 Gaucher C, Walrant-Debray O, Nguyen T-M, *et al*. PHEX analysis in 118 pedigrees reveals new genetic clues in hypophosphatemic rickets. Hum Genet 2009;125:401–11. 10.1007/s00439-009-0631-z
22 Women's and Children's Hospital. Women's and Children's Hospital; Women's and Children's Hospital. NewbornsInSA Research Study Adelaide, Australia, 2024. Available: https://www.wch.sa.gov.au/research/newbornsinsa-research-study [Accessed 8 Aug 2024].
23 Green RC, Shah N, Genetti CA, *et al*. Actionability of unanticipated monogenic disease risks in newborn genomic screening: Findings from the BabySeq Project. Am J Hum Genet 2023;110:1034–45. 10.1016/j.ajhg.2023.05.007
24 Green RC, Shah N, Genetti CA, *et al*. Medical evaluation of unanticipated monogenic disease risks identified through newborn genomic screening: findings from the babyseq project. Genetic and Genomic Medicine 2022. 10.1101/2022.03.18.22272284
25 Ceyhan-Birsoy O, Murry JB, Machini K, *et al*. Interpretation of Genomic Sequencing Results in Healthy and Ill Newborns: Results from the BabySeq Project. Am J Hum Genet 2019;104:76–93. 10.1016/j.ajhg.2018.11.016
26 Hao C, Guo R, Hu X, *et al*. Newborn screening with targeted sequencing: a multicenter investigation and a pilot clinical study in China. J Genet Genomics 2022;49:13–9. 10.1016/j.jgg.2021.08.008
27 Yang R-L, Qian G-L, Wu D-W, *et al*. A multicenter prospective study of next-generation sequencing-based newborn screening for monogenic genetic diseases in China. World J Pediatr 2023;19:663–73. 10.1007/s12519-022-00670-x
28 Bodian DL, Klein E, Iyer RK, *et al*. Utility of whole-genome sequencing for detection of newborn screening disorders in a population cohort of 1,696 neonates. Genet Med 2016;18:221–30. 10.1038/gim.2015.111
29 Chen T, Fan C, Huang Y, *et al*. Genomic Sequencing as a First-Tier Screening Test and Outcomes of Newborn Screening. JAMA Netw Open 2023;6:e2331162. 10.1001/jamanetworkopen.2023.31162
30 Huang X, Wu D, Zhu L, *et al*. Application of a next-generation sequencing (NGS) panel in newborn screening efficiently identifies inborn disorders of neonates. Orphanet J Rare Dis 2022;17:66. 10.1186/s13023-022-02231-x
31 Jian M, Wang X, Sui Y, *et al*. A pilot study of assessing whole genome sequencing in newborn screening in unselected children in China. Clin Transl Med 2022;12:e843. 10.1002/ctm2.843
32 Pavey AR, Bodian DL, Vilboux T, *et al*. Utilization of genomic sequencing for population screening of immunodeficiencies in the newborn. Genet Med 2017;19:1367–75. 10.1038/gim.2017.57
33 Balciuniene J, Liu R, Bean L, *et al*. At-Risk Genomic Findings for Pediatric-Onset Disorders From Genome Sequencing vs Medically Actionable Gene Panel in Proactive Screening of Newborns and Children. JAMA Netw Open 2023;6:e2326445. 10.1001/jamanetworkopen.2023.26445
34 Roman TS, Crowley SB, Roche MI, *et al*. Genomic Sequencing for Newborn Screening: Results of the NC NEXUS Project. Am J Hum Genet 2020;107:596–611. 10.1016/j.ajhg.2020.08.001
35 Cai L, Liu Y, Xu Y, *et al*. Multi-Center in-Depth Screening of Neonatal Deafness Genes: Zhejiang, China. Front Genet 2021;12:637096. 10.3389/fgene.2021.637096
36 Ma Z, Huang W, Xu J, *et al*. Analysis of deafness susceptibility gene of neonates in northern Guangdong, China. Sci Rep 2024;14:362. 10.1038/s41598-023-49530-2
37 Ye L, Yin Y, Chen M, *et al*. Combined genetic screening and traditional newborn screening to improve the screening efficiency of congenital hypothyroidism. Front Pediatr 2023;11:1185802. 10.3389/fped.2023.1185802
38 Luo H, Yang Y, Wang X, *et al*. Concurrent newborn hearing and genetic screening of common hearing loss variants with bloodspot-based targeted next generation sequencing in Jiangxi province. Front Pediatr 2022;10:1020519. 10.3389/fped.2022.1020519
39 Ariceta G, Beck-Nielsen SS, Boot AM, *et al*. The International X-Linked Hypophosphatemia (XLH) Registry: first interim analysis of baseline demographic, genetic and clinical data. Orphanet J Rare Dis 2023;18:304. 10.1186/s13023-023-02882-4
40 Jackson L, Weedon MN, Green HD, *et al*. Influence of family history on penetrance of hereditary cancers in a population setting. EClinicalMedicine 2023;64:102159. 10.1016/j.eclinm.2023.102159
41 Downie L, Bouffler SE, Amor DJ, *et al*. Gene selection for genomic newborn screening: Moving toward consensus? Genet Med 2024;26:101077. 10.1016/j.gim.2024.101077
42 Minten T, Bick S, Adelson S, *et al*. Data-driven consideration of genetic disorders for global genomic newborn screening programs. medRxiv 2025.:2024.03.24.24304797. 10.1101/2024.03.24.24304797
43 Beaumont RN, Wright CF. Estimating diagnostic noise in panel-based genomic analysis. Genet Med 2022;24:2042–50. 10.1016/j.gim.2022.06.008
44 Kim DeLuca E, Wu AC, Christensen KD, *et al*. Modernizing Newborn Screening in the Genomic Era: Importance of Health-Related Quality of Life. Pharmacoecon Open 2024;8:787–92. 10.1007/s41669-024-00528-0
45 Turnbull C, Firth HV, Wilkie AOM, *et al*. Population screening requires robust evidence—genomics is no exception. The Lancet 2024;403:583–6. 10.1016/S0140-6736(23)02295-X
46 Horton R, Wright CF, Firth HV, *et al*. Challenges of using whole genome sequencing in population newborn screening. BMJ 2024;384:e077060. 10.1136/bmj-2023-077060
47 Clark CCA, Boardman FK. Expanding the notion of "benefit": comparing public, parent, and professional attitudes towards whole genome sequencing in newborns. New Genet Soc 2022;41:96–115. 10.1080/14636778.2022.2091533

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/bmjmed-2025-001726).