



Article

A Delphi Survey Study to Formulate Statements on the Treatability of Inherited Metabolic Disorders to Decide on Eligibility for Newborn Screening

Abigail Veldman ¹, M. B. Gea Kiewiet ², Dineke Westra ³, Annet M. Bosch ⁴, Marion M. G. Brands ⁴, René I. F. M. de Coö ⁵, Terry G. J. Derks ¹, Sabine A. Fuchs ⁶, Johanna. M. P. van den Hout ⁷, Hidde H. Huidekoper ⁷, Leo A. J. Kluijtmans ⁸, Klaas Koop ⁶, Charlotte M. A. Lubout ¹, Margaretha F. Mulder ⁴, Bianca Panis ⁹, M. Estela Rubio-Gozalbo ¹⁰, Monique G. de Sain-van der Velden ¹¹, Jaqueline Schaefer ⁹, Andrea B. Schreuder ¹, Gepke Visser ^{4,6}, Ron A. Wevers ⁸, Frits A. Wijburg ⁴, M. Rebecca Heiner-Fokkema ¹² and Francjan J. van Spronsen ^{1,*}

- ¹ Division of Metabolic Diseases, Beatrix Children's Hospital, University of Groningen, University Medical Center Groningen, 9718 GZ Groningen, The Netherlands
 - ² Department of Genetics, University of Groningen, University Medical Center Groningen, 9718 GZ Groningen, The Netherlands
 - ³ Department of Human Genetics, Radboud University Medical Center, 6525 GA Nijmegen, The Netherlands
 - ⁴ Department of Pediatrics, Division of Metabolic Disorders, Emma Children's Hospital, Amsterdam University Medical Centre, 1105 AZ Amsterdam, The Netherlands
 - ⁵ Department of Toxicogenomics, Unit Clinical Genomics, MHeNs School for Mental Health and Neuroscience, Maastricht University, 6229 ER Maastricht, The Netherlands; rene.decoo@maastrichtuniversity.nl
 - ⁶ Department of Metabolic Diseases, University Medical Center Utrecht, Wilhelmina Children's Hospital, 3584 EA Utrecht, The Netherlands
 - ⁷ Department of Pediatrics, Center for Lysosomal and Metabolic Diseases, Erasmus University Medical Center, 3015 GD Rotterdam, The Netherlands
 - ⁸ Department of Human Genetics, Translational Metabolic Laboratory, Radboud University Medical Center, 6525 GA Nijmegen, The Netherlands; ron.wevers@radboudumc.nl (R.A.W.)
 - ⁹ Department of Pediatrics, Maastricht University Medical Center, 6229 HX Maastricht, The Netherlands
 - ¹⁰ Department of Pediatrics and Clinical Genetics, Maastricht University Medical Center, 6229 HX Maastricht, The Netherlands
 - ¹¹ Section Metabolic Diagnostics, Department of Genetics, University Medical Center Utrecht, 3584 CX Utrecht, The Netherlands
 - ¹² Department of Laboratory Medicine, Laboratory of Metabolic Diseases, University of Groningen, University Medical Center Groningen, 9718 GZ Groningen, The Netherlands
- * Correspondence: f.j.van.spronsen@umcg.nl



Citation: Veldman, A.; Kiewiet, M.B.G.; Westra, D.; Bosch, A.M.; Brands, M.M.G.; de Coö, R.I.F.M.; Derks, T.G.J.; Fuchs, S.A.; van den Hout, J.M.P.; Huidekoper, H.H.; et al. A Delphi Survey Study to Formulate Statements on the Treatability of Inherited Metabolic Disorders to Decide on Eligibility for Newborn Screening. *Int. J. Neonatal Screen.* **2023**, *9*, 56. <https://doi.org/10.3390/ijns9040056>

Academic Editor: Jim R. Bonham

Received: 10 September 2023

Revised: 4 October 2023

Accepted: 5 October 2023

Published: 11 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: The Wilson and Jungner (W&J) and Andermann criteria are meant to help select diseases eligible for population-based screening. With the introduction of next-generation sequencing (NGS) methods for newborn screening (NBS), more inherited metabolic diseases (IMDs) can technically be included, and a revision of the criteria was attempted. This study aimed to formulate statements and investigate whether those statements could elaborate on the criterion of *treatability* for IMDs to decide on eligibility for NBS. An online Delphi study was started among a panel of Dutch IMD experts (EPs). EPs evaluated, amended, and approved statements on *treatability* that were subsequently applied to 10 IMDs. After two rounds of Delphi, consensus was reached on 10 statements. Application of these statements selected 5 out of 10 IMDs proposed for this study as eligible for NBS, including 3 IMDs in the current Dutch NBS. The statement: 'The expected benefit/burden ratio of early treatment is positive and results in a significant health outcome' contributed most to decision-making. Our Delphi study resulted in 10 statements that can help to decide on eligibility for inclusion in NBS based on *treatability*, also showing that other criteria could be handled in a comparable way. Validation of the statements is required before these can be applied as guidance to authorities.

Keywords: newborn screening; treatability; treatable; Wilson and Junger criteria; inborn errors of metabolism; inherited metabolic disorders; next-generation sequencing

1. Introduction

Newborn screening (NBS) aims to identify disorders early to prevent or significantly reduce morbidity and mortality. To guide the selection of disorders that could qualify for population-based screening, 10 criteria were established by Wilson and Junger (W&J) in 1968 [1]. These W&J criteria and their revised version by the World Health Organization (WHO) in 2008 and 2011 [2,3] have been the gold standard for the principles for public health screening policies ever since and are used to decide on the inclusion of disorders in NBS [1–4], despite the fact that they were not originally meant for NBS.

Worldwide, the procedure to include disorders in NBS programs and the number of diseases included in NBS programs varies [5,6]. The United States of America and Denmark use the quantitative scoring matrix of the American College of Medical Genetics (ACMG) to decide on the inclusion of disorders for NBS [6]. But even this Recommended Universal Screening Panel (RUSP) [7,8] has not prevented these countries from selecting different disorders [9]. Australia uses its own National Policy Framework Newborn Blood Spot Screening based on the criteria of W&J, which are “amended to suit NBS and the local context in which programs operate” [10,11]. This National Policy framework references the use of RUSP and the governing documents of the United Kingdom and New Zealand. Still, there are national and even regional differences in the selection of disorders screened for [12]. For example, the state of Victoria decided not to screen for galactosemias, while other states do [12]. These differences can partly be explained by the fact that in practice, the choices regarding disorders included in screening programs often depend on financial support, technical and medical knowledge, and/or the personal interest of healthcare professionals, scientists, and other persons involved in policy-making [5,13]. Despite the guidance of the W&J criteria, there is room for different interpretations regarding the choice to screen for disorders. Unfortunately, it is often not clear on what basis a disease has or has not been selected.

The rationale behind selecting disorders becomes increasingly relevant since we are at the beginning of a new era: a genetic-based NBS [9,14–19]. Next-generation sequencing (NGS) techniques allow us to include a much higher number of monogenetic disorders compared to any existing NBS program, albeit with unknown sensitivity. To be able to be seen as a suitable test for NBS in general, these techniques need to pass technical, ethical, and financial hurdles first. If these methods prove suitable, we will need to discuss the other W&J for each disorder. One of the most important W&J criteria in this respect is probably: ‘There should be an accepted treatment for patients with recognized disease’. Within the context of genetic screening possibilities, Andermann et al. [2,3] developed an approach to guide genetic screening policy-making. Twenty criteria were designed, including criterion 17 on intervention: “There should be an accepted intervention (ex. prevention, treatment, family planning) that forms part of a coherent management system”. The aspect of an accepted treatment or *treatability* consists of three main requisites: (1) the presence of treatment for this disorder, (2) the approval of this treatment by the FDA/EMA, and (3) financial coverage or reimbursement by standard healthcare. Without *treatability*, no disorder will be approved to be selected for NBS [2,3].

Van Karnebeek et al. reported on the amenability of inherited metabolic disorders (IMDs) in patients presenting with epilepsy or mental retardation [20,21]. However, it may be questioned whether these IMDs meet the criteria on *treatability* in the context of NBS. The literature lacks a clear definition of *treatability* in the context of NBS. Also, recent innovations, like NGS techniques and improved treatment possibilities for many disorders, generate new ideas about what is a treatable disease. To make an attempt to ease the selection process for diseases potentially eligible for NBS, we aimed to create a list of statements in the context

of NBS that elaborates on *treatability* using a Delphi approach [22]. Since the current NBS mainly consists of IMDs, we took this group of disorders as an example.

2. Materials and Methods

The methodology of our project was based on the Recommendations for the Conducting and REporting of DELphi Studies (CREDES) [22] stated by EQUATOR network.org (accessed on 3 April 2022), and the further available literature reporting on Delphi Studies [23–26]. The rationale behind the Delphi technique builds on the assumption that the opinion of a group is more valid than an individual opinion [27,28]. This is used to form a consensus or to explore a field beyond the borders of the current knowledge and conceptual world [22], for example, the term “*treatability*”. Different from the Nominal Group Technique or a consensus conference [27], the Delphi technique is an anonymous process in which every opinion will be heard, preventing the dominance of more influential experts in the discussion. Moreover, it provides the possibility to think about answers. Differently from an interview or a regular survey, the constructive nature of the Delphi technique allows for the generation of new ideas, the possibility to respond to other experts and the chance to re-state one’s opinion [22,29], which seems perfect for a medical ethical discussion on the criterion of *treatability* in the context of NBS.

2.1. The General Design of the Delphi Study

Our Delphi study was designed and conducted in three anonymous online survey rounds (Part 1.1, Part 1.2, and Part 2). The aim of Part 1 was to develop statements and reach consensus on these statements that together could define *treatability* in NBS. The aim of Part 2 was to test these statements on 10 IMDs, to gain insight into which statement contributed the most, and investigate whether it is possible to put weight on each statement. Figure 1 illustrates a flowchart of the design of the study.

2.1.1. Part 1

In Part 1.1, the research team (RT: AV, MRH, FJvS), supported by two medical ethicists (EM, WD), formulated 28 statements subcategorized into seven items (Appendix A) based on a literature review on *treatability*. Statements were also inspired by the NEXUS study [17]. The professionals invited to participate in this study were colleagues with known experience in IMDs from the Dutch Advisory Committee Neonatal Screening for IMDs and the project group members involved in the study on NGS as a first-tier approach in NBS in the Netherlands (NGSf4NBS) [18]. Together, they formed the Expert Panel (EP). Invitations were sent by email. The EP consisted of pediatricians for IMDs (N = 17), clinical laboratory geneticists/chemists (N = 8), clinical laboratory geneticists/chemists (N = 1), and pediatric neurologists with experience with IMDs (N = 2). This total of 28 EP members (EPs) were asked to rate the quality of the *treatability* statements on a 1–10 Likert scale (1 = completely disagree, 10 = strongly agree) and to evaluate whether this statement should be added to the final decision matrix, or not. It was encouraged to substantiate choices by commenting on the statements or to add new relevant statements in special open-field comment boxes. In Part 1.2, all results were anonymized and shared via email. Some of the statements could be improved according to the comments made by the EPs in Part 1.1. Again, the EPs were asked to rate every statement on a 1–10 Likert scale and to comment on and improve the statements. EPs were given a response time of one week each for Part 1.1 and Part 1.2. Before the start of the study, the RT decided that statements met consensus if they had a mean of at least 7.0 and a median of at least 7.0, and a mode of at least 7.0. Statements that met consensus proceeded to Part 2.

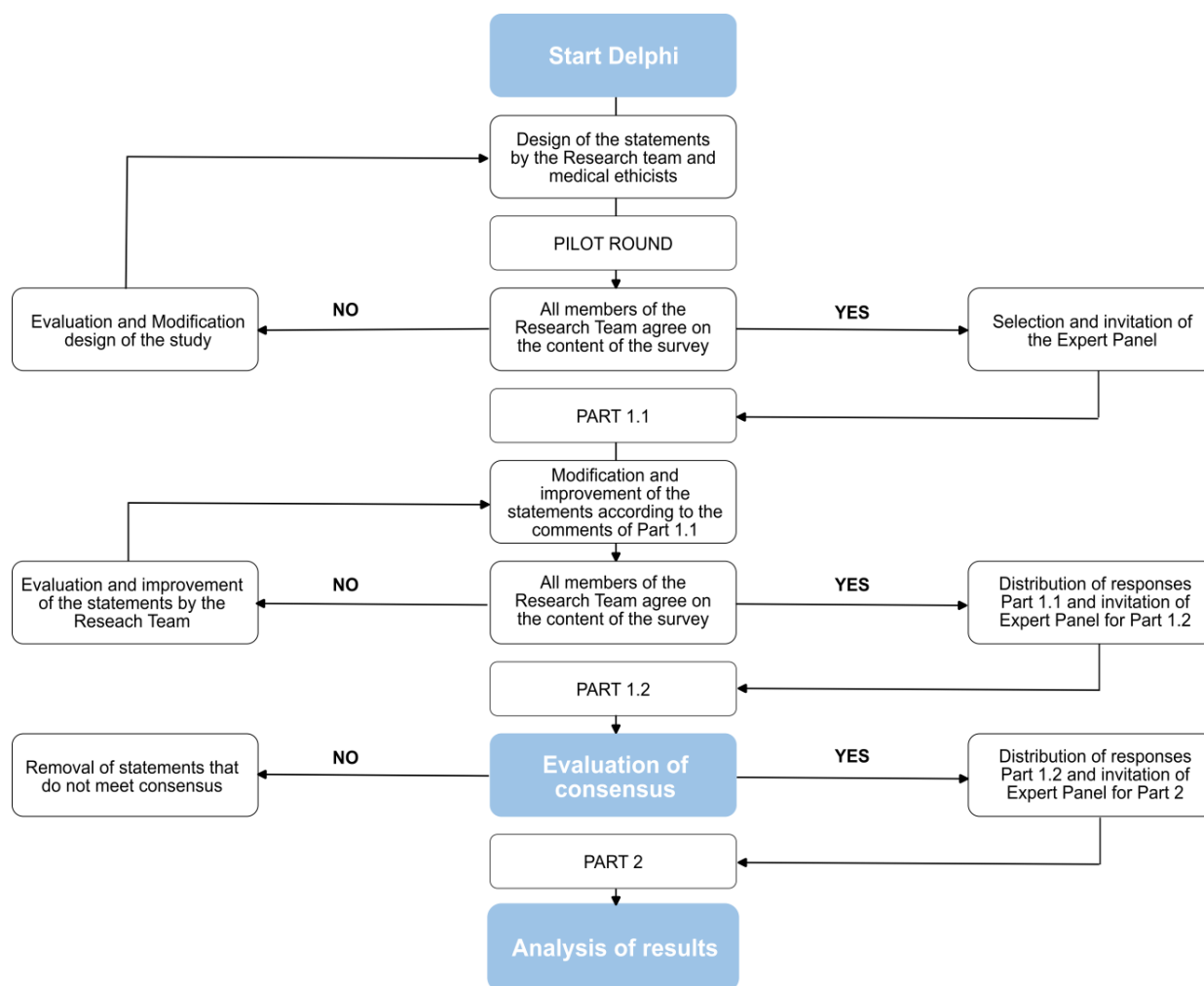


Figure 1. Flowchart of the design of the Treatability Delphi study.

2.1.2. Part 2

In Part 2, EPs were asked to evaluate 10 IMDs on their eligibility for NBS using the statements that met the consensus criterion in Part 1. The selection of these 10 IMDs as examples were based on different IMDs selected by the North Carolina Newborn Exome Sequencing for Universal Screening (NEXUS), The BabySeq Project, RUSP, and our selection of IMDs made in the NGSf4NBS project [18]. We chose seven IMDs that were not selected in each of the studies and for which *treatability* may be debatable. Three IMDs already included in the Dutch NBS served as controls. The 10 IMDs selected for this study were: pyridoxine-dependent epilepsy (PDE; OMIM 266100), classic galactosemia (CG; OMIM 606999), carnitine palmitoyltransferase 2 deficiency (CPT2; OMIM 255110, 600649, 608836), glycogen storage disease type 2 (Pompe disease; GSD2; OMIM 232300), autosomal recessive guanosine triphosphate cyclohydrolase 1 deficiency (GCH1; OMIM 233910), ornithine transcarbamylase deficiency (OTC; OMIM 311250), Wilson’s Disease (WD; OMIM 277900), methylmalonic aciduria due to methyl malonyl-CoA mutase deficiency (MCM; OMIM 251000), tyrosine hydroxylase deficiency (TH; OMIM 605407), and phenylketonuria (PKU; OMIM 261600). For each IMD, three questions were asked: First, to score the IMD on each consented statement on a 0–5 Likert scale (0 not applicable/true, and 5 most applicable/true). Next, to score on a 0–5 Likert scale if the IMD should be included in NBS (0 do not agree, 5 completely agree). An open-field box was added to give EPs the option to elaborate on their views. Finally, the EPs were asked to discriminate between the importance of statements for the final decision matrix. This was achieved via a multiple-choice question to rank which statement(s) contributed most to the decision

to include, or not include, this IMD in the NBS. The options were: ‘Did not contribute to my decision at all’ (0 points), ‘Small contribution but should be in the matrix’ (1 point), ‘moderate contribution’ (2 points), and ‘Large contribution (3 points).’ In case an IMD clinically presents with multiple phenotypes, they were asked to reason from the most severe phenotype. For Part 2, the EPs had two weeks to fill in the survey.

2.2. Informational Input and Piloting of Materials

Qualtrics XM, a web-based survey tool, was used (Qualtrics, Seattle 2002). Qualtrics XM (version March 2022) allows for the design, distribution, and processing of online surveys in multiple formats with a professional layout. For every round, a separate survey was created. Therefore, in total, three surveys were created. All responses were anonymous and EPs received an invitation by email with an anonymous link. Every methodological step was thoroughly evaluated by the RT and piloted in advance to test the quality, accessibility, and user-friendliness.

2.3. Strategy for Interpretation and Processing of Results

Between each round, a report of all results was constructed by one of the members of the research team (AV). This report consisted of statistical data, including the mean, median, and mode of all statements, and qualitative comments of the EP per statement and item. All results, including partial responses, were discussed within the RT and the consensus, and decision rules were applied (Table 1). After this, the survey of the new round was created, evaluated, and piloted again within the RT. After approval by all three members of the RT, the Qualtrics XM link for the new round was distributed, together with the result report available for inspection by the EPs via email.

Table 1. Consensus and decision rules after each round.

| Part 1.1/Round 1 | Part 1.2/Round 2 | Part 2/Round 3 |
|---|--|--|
| Consensus was declared if a statement scored a mean of ≥ 7.0 , a median of ≥ 7.0 , and a mode of ≥ 7.0 . Regardless of whether consensus was reached, statements proceeded to Part 1.2 to be discussed in an altered form. | Consensus was declared if a statement scored a mean of ≥ 7.0 , a median of ≥ 7.0 , and a mode of ≥ 7.0 . Only statements reaching consensus proceeded to Part 2 ¹ . | A $\geq 75\%$ majority agreement by the EP had to be reached to include a disorder in the NBS based only on <i>treatability</i> . |
| The RT was free to add small grammatical or textual changes to clarify the statement (also based on individual comments from the EP), provided that it did not compromise the meaning or implication of the statement. | The RT was free to add small grammatical or textual changes to clarify the statement (also based on individual comments from the EP), provided that it did not compromise the meaning or implication of the statement. | The RT was free to add small grammatical or textual changes to clarify the statement (also based on individual comments from the EP), provided that it did not compromise the meaning or implication of the statement. |
| If 3 or more EPs requested the same alteration, the statement was adapted accordingly. | No new statements were added after Part 1.2. | |
| If 3 or more EPs found the statement irrelevant, it could be removed, adapted, specified, or combined with another statement. | | |
| The RT was free to add statements, based on the suggestions of individual panelists. | | |

¹ An exception was made for Q6 and Q7 (Appendix A), in which only the statements that reached consensus that indicated the minimum number of publications or number of patients were added to the final list of statements. EP = expert panel, EPs = expert panel members, IMDs = inherited metabolic disorders, NBS = newborn screening, RT = research team.

Part 2 required a strategy for the interpretation of the results and translation into a final decision matrix. First, the EPs looked at the contribution of the statements on a 3-point Likert scale: no contribution (0 points), small contribution (1 point), moderate contribution (2 points), and large contribution (3 points). The statements that contributed most to all IMDs were regarded as important in decision-making and received a higher weight in the final list. Second, we analyzed if a disorder should be included in the NBS or not according to the vast majority ($>75\%$) of EPs. Last, we assessed if there was a correlation between

an IMD with a high score on the *treatability* statements and an agreement on eligibility for NBS for that IMD.

3. Results

3.1. Results Part 1.1 and Part 1.2

3.1.1. Panel Participation

Twenty-eight out of forty-one (68.3%) invited Eps filled in the Delphi survey and formed the EP in Table 2. Since the survey was anonymous, the EPs were asked to confirm their participation via email, so that we could send targeted emails for the second and third rounds. Five of them failed to do so. Therefore, only 23 EPs (56%) were invited to participate in Part 1.2 and 2. Of them, 21 EPs (91%) participated in Part 1.2.

Table 2. Final statements defining *treatability* for inherited metabolic disorders.

| | Statement (S)/Question (Q) |
|-----------|---|
| Q1./S1 | There is a treatment available that is fully financially covered or reimbursed by standard health care (also when a patient reaches adulthood). |
| Q2./S2 | The expected benefit/burden ratio of early treatment is positive and results in significant health benefits. |
| Q3./S3 | The treatment/diet results in enough significant health benefits (compared to no treatment) to accept a small risk for disease-related mortality. |
| Q5.1./S4 | Early detection with subsequent treatment, when compared to clinical presentation with subsequent treatment, prevents (sudden) death. |
| Q5.4./S5 | Early detection with subsequent treatment, when compared to clinical presentation with subsequent treatment, prevents symptoms clearly related to one or two primary organ(s). |
| Q5.5./S6 | Early detection with subsequent treatment, when compared to clinical presentation with subsequent treatment, prevents developmental delay. |
| Q5.7./S7 | Early detection with subsequent treatment, when compared to clinical presentation with subsequent treatment, improves quality of life. |
| Q6.4./S8 | Papers on treatments were published by at least two institutes with good quality data about a plausible mechanism with an “adequate number of patients” with clear effect size and suggesting the efficacy of early treatment is sufficient to accept that treatment for newborn screening. |
| Q6.5./S9 | There is consensus on positive outcome of early detection by an (international) expert meeting. |
| Q7.3./S10 | The expected effect of treatment was demonstrated in more than 75% of patients (both mild and severe variants). |

3.1.2. Statements and Item Throughput

Part 1.1 started with seven main items concerning the treatability of disorders. The seven main items included 26 statements on: (Q1) financial reimbursement by insurance/EMA approval; (Q2) the accepted level of burden or risk for the newborn/child related to the available treatment; (Q3) the risk of disease-related mortality despite treatment; (Q4) the risk of disease-related morbidity despite treatment; (Q5) the outcomes should at least be prevented, reversed, or improved by early detection with consequent treatment through NBS; (Q6) the minimum amount of publications in which this outcome should have been demonstrated; (Q7) the minimum number of patients (both mild and severe variants) in which (any) effect of treatment has been demonstrated, (Appendix A). After ranking the statements in the first round, a total of N= 99 comments was made on the items and statements. At the same time, N = 23 suggestions on the “minimum number of publications” in item Q6.5 or “minimum number of patients” in items Q7.5 were made (Appendix A). After Part 1.1, 16 statements would have reached preliminary consensus based on our criteria. However, for many of them, major adaptations and specifications were requested as side comments. After processing and interpreting the results of Part 1.1, we followed the decision rules of Table 1. Based on that, 9 of 26 statements were removed (two of them were combined with another statement), 5 new statements were added and 7 statements were adapted or specified, while 9 statements were left unchanged, resulting

in a total of 22 statements at the end of Part 1.1 (Figure 2). The improved survey was the starting point of Part 1.2.

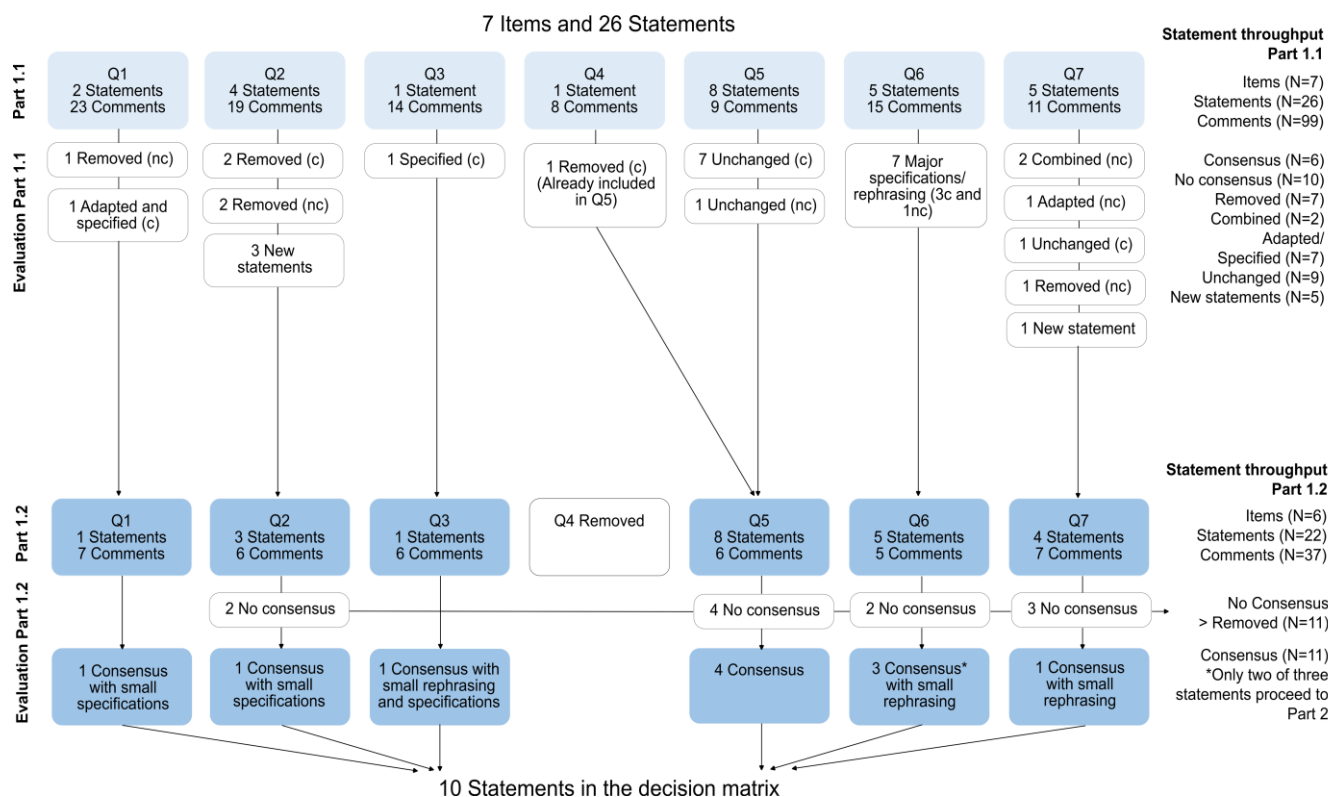


Figure 2. Item and statement throughput of the Delphi process in Part 1.1 and Part 1.2. Q = question/item, nc = no consensus, c = consensus * An exception was made for Q6 and Q7 (Appendix A) in which only the statements that indicate “the minimum number of publications” or “minimum number of patients” and reached consensus were added to the final list of statements.

Part 1.2 started with six questions/items, which were divided into a total of 22 statements (Appendix B). Considerably fewer comments (N = 37) were recorded after this round. Ten statements (45%) reached consensus and proceeded to Part 2. Minor grammatical specifications and rephrasing were needed on the 10 consented statements after Part 1.2. All items and statements of the first two rounds and their alterations can be found in Appendices A and B. The final list of statements is depicted in Table 2.

3.2. Results Part 2

3.2.1. Panel Participation

Twenty-two of twenty-three EPs (95.7%) filled in the survey of Part 2. However, some respondents failed to complete their responses for all 10 proposed IMD, especially due to a self-noted feeling of inadequate knowledge of specific IMDs and due to unknown reasons. The mean number of responses per statement was N = 16.3 (range 11–22) of the total of 23 responses.

3.2.2. Contribution of Statements

EPs were asked to rate per IMD which statement contributed most to their decision to include a disorder in the NBS or not based only on its *treatability*. The mean contribution score of every statement ranged from 1.8 to 2.6 points (minimum 0.0 and maximum 3.0). Statements 2 (2.6 points), 5 (2.4 points), and 7 (2.4 points) contributed more than the other statements. Statement 10 (1.8 points) contributed the least to the decision-making according to the EP.

3.2.3. Inclusion in NBS

Table 3 presents an overview of the selected IMDs used to test our 10 statements and the decision to consider these IMDs eligible or not for NBS based on the criterion of *treatability*. This Table also shows whether the IMDs were included in the panels of RUSP [8], and the studies of NEXUS [17], and The BabySeq Project [16,30,31]. According to >75% of the 22 responding EPs, 5 out of 10 IMDs would be added to the NBS, including the 3 disorders that are in the current Dutch NBS (PKU, MCM, and CG). Comments made regarding decision-making in Part 2 are depicted in Appendix C.

Table 3. Ten selected inborn metabolic disorders and assessment of *treatability* in different studies.

| | Name of Disorder | Associated Gene | OMIM ² | Assessment of Genes in Studies (Yes/No) * | | | | Treatability Delphi Study |
|--------|------------------|-----------------|-------------------|---|----------|------------|----------------------|---------------------------|
| | | | | NGSf4NBS Project [18] | RUSP [7] | NEXUS [17] | Babyseq [16] Project | |
| IMD 1 | PDE | ALDH7A1 | 266100 | Yes | No | No | No | Yes |
| IMD 2 | CG | GALT | 606999 | Yes | Yes | Yes | Yes | Yes |
| | | | 255110 | | | | | |
| IMD 3 | CPT2 | CPT2 | 600649 | | | | No | No |
| | | | 608836 | | | | | |
| IMD 4 | GSD2 | GAA | 232300 | No | Yes | Yes | Yes | No |
| IMD 5 | GCH1 | GCH1 | 233910 | Yes | Yes | Yes | No | No |
| IMD 6 | OTC | OTC | 311250 | Yes | No | Yes | Yes | No |
| IMD 7 | WD | ATP7B | 277900 | Yes | No | No | Yes | No |
| IMD 8 | MCM | MMUT | 251000 | Yes | Yes | No | No | Yes |
| IMD 9 | TH | TH | 605407 | Yes | No | No | No | Yes |
| IMD 10 | PKU | PAH | 261600 | Yes | Yes | Yes | Yes | Yes |

* Yes/No: These genes were/were not included in similar studies, listed in this table, in which eligibility and genetic screening for newborn screening was tested. ² OMIM = Online Mendelian Inheritance in Man. Every inborn metabolic disorder is identifiable by an OMIM code that can be used in databases. IMD = inherited metabolic disorder, NGSf4NBS = Next-generation Strategies First for Newborn Screening, RUSP = Recommended Uniform Screening Panel, NEXUS = North Carolina Newborn Exome Sequencing for Universal Screening study, Babyseq = The Babyseq Project. PDE = pyridoxine-dependent epilepsy, CG = classic galactosemia, CPT2 = carnitine palmitoyltransferase 2 deficiency, GSD 2 = glycogen storage disease type 2, GCH1 = autosomal recessive guanosine triphosphate cyclohydrolase 1 deficiency, OTC = ornithine transcarbamylase deficiency, WD = Wilson's Disease, MCM = methylmalonic aciduria due to methyl malonyl-CoA mutase deficiency, TH = tyrosine hydroxylase deficiency, PKU = phenylketonuria.

3.2.4. The Scoring of Statements per IMD

To assess whether the selection of IMDs for NBS, as described in 6.2.2, correlates with a high score on the *treatability* statements, all mean scores on the 1–5 Likert scale were calculated in Table 4. The first observation was that both PKU and CG, which are already included in the Dutch NBS, scored the highest on almost every statement, resulting in mean scores above 4.0.

For the other IMDs, there was no direct relationship between scoring high on the *treatability* statements and their eligibility for NBS. For example, GSD2 and OTC both scored higher than PDE, MCM, and TH, but were not selected by the EP as eligible for NBS. Additionally, the opinion about the level of *treatability* of some IMDs was strongly divided among respondents. Table S5 in the Supplementary Materials presents the variation in the responses of the EPs.

Table 4. Mean ranking scores ¹ of the *treatability* statements per inherited metabolic disorder.

| | PKU | CG | MCM | PDE | TH | CPT2 | WD | GCH1 | GSD2 | OTC |
|--|-------------------|------------------|------------------|------------------|------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Consensus on eligibility for NBS by the EP | Yes (−100.00%) | Yes (−94.70%) | Yes (−87.60%) | Yes (−86.40%) | Yes (−80.00%) | No (−72.20%) | No (−71.20%) | No (−66.70%) | No (−58.80%) | No (−50.00%) |
| Response rate | N = 16 | N = 19 | N = 16 | N = 22 | N = 15 | N = 18 | N = 14 | N = 12 | N = 17 | N = 16 |
| S1 | 4.75 | 4.78 | 4.5 | 3.9 | 4.56 | 4.38 | 4.27 | 4.23 | 4.24 | 4.44 |
| S2 | 4.88 | 4.42 | 3.94 | 4.05 | 4.19 | 3.88 | 3.93 | 4.08 | 4.06 | 3.63 |
| S3 | 4.43 | 4.26 | 3.63 | 4.14 | 3.67 | 3.39 | 4.07 | 3.55 | 4.12 | 4 |
| S4 | 2.79 | 4.22 | 3.69 | 3.4 | 2.87 | 3.56 | 2.92 | 2.33 | 3.59 | 3.94 |
| S5 | 4.75 | 4.35 | 3.56 | 3.86 | 3.69 | 3.78 | 4.07 | 3.77 | 4.24 | 3.88 |
| S6 | 4.75 | 3 | 3.56 | 3.43 | 3.69 | 3 | 3.83 | 3.38 | 2.88 | 3.44 |
| S7 | 4.56 | 3.89 | 3.6 | 4.1 | 4 | 3.44 | 4.21 | 3.85 | 4.12 | 3.69 |
| S8 | 4.88 | 4.32 | 3.75 | 4 | 3.21 | 3.22 | 3.54 | 2.67 | 4.12 | 3.75 |
| S9 | 4.88 | 4.21 | 3.53 | 3.76 | 2.87 | 3 | 3.46 | 2.92 | 4.07 | 3.56 |
| S10 | 4.75 | 3.95 | 3.13 | 3.9 | 3.21 | 3 | 3.21 | 3.17 | 3.56 | 3.38 |
| Mean ² per IMD | 4.54 | 4.14 | 3.69 | 3.85 | 3.6 | 3.46 | 3.75 | 3.39 | 3.9 | 3.77 |

¹ All scores are based on a 1–5 score on a Likert scale. ² The mean per IMD is calculated from the means of all statements. IMD = inherited metabolic disorder, NBS = newborn screening, EP = expert panel, S = statement, PDE = pyridoxine-dependent epilepsy, CG = classic galactosemia, CPT2 = carnitine palmitoyltransferase 2 deficiency, GSD 2 = glycogen storage disease type 2. GCH1 = autosomal recessive guanosine triphosphate cyclohydrolase 1 deficiency, OTC = ornithine transcarbamylase deficiency, WD = Wilson’s Disease, MCM = methylmalonic aciduria due to methyl malonyl-CoA mutase deficiency, TH = tyrosine hydroxylase deficiency, PKU = phenylketonuria.

3.2.5. Comments and Evaluation of the EP on the Decision Matrix

The EP was asked an optional open question to give their opinion on whether the statements are a useful tool to quantify *treatability* for individual IMDs in the context of NBS. Twelve of twenty-two EPs (54.5%) responded. Responses varied and could be classified into five main opinions:

- Opinion 1: It can be a useful tool if all data are available and experts are involved in decision-making.
- Opinion 2: It can be a useful tool as a starting point for a discussion for inclusion in NBS.
- Opinion 3: Concerns about how to proceed with this tool if disorders are not eligible for NBS.
- Opinion 4: Concerns about how to proceed with this tool for disorders with a broad phenotypic variability.
- Opinion 5: The decision matrix is not useful (yet) to quantify *treatability* for IMDs in NBS.

A complete overview of all responses can be found in Appendix D.

4. Discussion

To the best of our knowledge, this is the first study that designed and performed an online Delphi study with pediatricians and experts on IMDs to formulate statements that help to elaborate on the definition of *treatability* of IMDs in the context of NBS. For this, we introduced a score for *treatability* to assess their eligibility for NBS. In this pioneering study, the statement: “The expected benefit/burden ratio of early treatment is positive and results in a significant health benefit” was found to contribute most clearly to decision-making. The final 10 statements on *treatability* show that a Delphi study with clear consensus and decision rules seems to be a suitable method to create a scoring system on *treatability*. This study also provides more insight into the aspects that IMD experts find important when deciding on the *treatability* of IMDs in the context of NBS, acknowledging the need for validation by international colleagues and patient/parent representatives.

Several methodological aspects should be taken into account before interpreting our results. First, the Delphi technique is one of the most valuable methods to reach consensus on research questions involving a medical ethical discussion [32] and is particularly interesting for topics without clear (data on) consensus [28,32]. This technique is based on expert opinions, which is considered to have the lowest level of evidence in evidence-based medicine according to systems for guideline development [33–36]. The validity and quality of a Delphi study, however, mainly depend on the design, conduction, analysis, and modification of results between rounds, and the reporting of the Delphi by the investigator,

rather than the technique itself [22,37]. By following CREDES [21], and ensuring complete and clear reporting on the entire study, we believe that this study is a first and important step towards a definition of the criterion of *treatability* for NBS. A risk of cognitive biases may occur if the Delphi study is not designed properly due to the formulation and the way in which the survey items and statements are presented [26,38]. We tried to follow the recommendations of Markmann et al., 2021 [38], to minimize the amount of abstract language use and information in surveys, by discussing both the content and the formulation of statement proposals with ethicists (EM, WD). We also tried to prevent framing and anchoring. Framing occurs when the EP is not sufficiently heterogeneous and group-thinking occurs when the EPs have the same background [39,40]. This may lead to a polarized judgment of the statements that is not representative. Framing was at least partly avoided by performing this study anonymously, while the group itself consisted of a wide variety of specialists involved in various aspects of the diagnosis and treatment of IMDs. In a study requiring such specific knowledge of IMDs, only possessed by a few specialists, more extensive heterogeneity is very hard to achieve. Anchoring may occur when the statements are presented in a certain manner or order. We could not easily randomize statements as they, at least in part, refer to each other. Some anchoring could thus not be prevented. Of course, the involvement of IMD professionals can also be considered a strength of this study. The EPs are highly experienced in delivering the best care for IMD patients and are up-to-date on the latest treatment options, which makes them adequate candidates to elaborate on the concept of *treatability*. Therefore, we think that our data are valid. At the same time, validation and weighting of the statements in other groups of professionals, e.g., international colleagues with experience in IMDs and medical ethicists, and patient representative organizations, is needed.

Our statements were primarily based on the W&J criterion 2 and part of the Andermann criterion 17 on *treatability*. No other criteria were considered. On the one hand, we think that each of the criteria deserve to be evaluated in a transparent process. On the other hand, our study shows that it is hard to evaluate each criterion on its own, as *treatability* is highly interrelated to other criteria, such as costs and phenotypic variability (e.g., age of onset, severity). The mean scores also do not take into account the weight of the contribution of the statements. Most information is contained in the details, including the outlier remarks for each statement. These issues could, at least in part, explain the differences seen between the lower-ranking *treatability* scores of PDE, MCM, and TH and the overall conclusion that these disorders are still eligible for NBS, or the opposite, i.e., a high score but not considered eligible for NBS, as in GSD2, OTC, and WD. Likewise, the value of a single statement in the criterion of *treatability* can be low, e.g., statement 6 in GSD2, whereas the statements as a whole provide a more complete impression of the level of *treatability*. The transparency of the process and a highly experienced EP team per IMD are, therefore, essential.

IMDs are (extremely) rare, resulting in a lack of data on their natural history and phenotypic variability, and not all EPs could obtain a complete overview of all IMDs. They may very well impact the EPs' ideas regarding the aspect of *treatability*, especially if these aspects are not yet addressed individually. The lack of knowledge of natural history and phenotypic variability could also explain the low response rate for some IMDs in Part 2. Between three and eight EPs (13.6–36.4%) failed to respond, stressing the fact that a panel of acknowledged disease experts for each disease is necessary to assess their eligibility for NBS. From this response, it can be concluded that the EPs felt they needed a substantial level of knowledge of an IMD to decide on its *treatability* or eligibility for NBS, knowledge that is limited in the case of rare disorders. Experts on specific IMDs and endocrinological diseases and medical ethicists are needed to ultimately decide on the *treatability* and eligibility of disorders for NBS. Such studies are also needed with parents (to be). This may seem in contrast to the high level of expertise EPs considered necessary to evaluate the statements, but the studies of Armstrong et al., 2022, from the Babyseq project, and the study of Blom

et al., 2021, provided at least some evidence that parents do understand the concept of the choice of whether to screen for (un)treatable disorders [41–43].

The necessity to obtain a group of disease-specific experts is exemplified by the outcomes of several IMDs. In MCM, scores were relatively low on statements 9 and 10. The discrepancy between agreement on eligibility for NBS and the *treatability* score of MCM might be the result of the differences in severity in the presentation of this IMD; EPs agreed that severe MCM is not fit for inclusion in NBS, while milder MCM variants were found to be eligible, showing that the statements are difficult to assess for IMD, with large phenotypic variances. In PDE, scores are relatively low on statement 4 and statement 6. However, this seems to be due to the opinion of one EP member, who was a clear outlier in the score assessments. This EP member opposed a different approach to this IMD by reasoning that PDE should be included in the guidelines for the treatment of epilepsy rather than including PDE in the NBS. This can also be learned from CPT2. In CPT2, for example, the EP considered more knowledge necessary to be able to judge the severity of disease presentation in order to reduce the potential risk of overtreatment in patients with less severe disease. Based on the comments of the EP, some statements deserve minor improvements, including the addition of a more detailed description of ‘significant health benefits’ (statement 2), the addition of a ‘not-applicable’ option to the statements (e.g., statement 3 is not applicable for PKU), and the addition of consensus on the positive outcome of early detection by an (international) expert meeting (statement 9).

In conclusion, we built a scoring system based on the statements. This Delphi study allowed us to gain insight into *treatability* in the context of NBS. Our study shows that, with solid statements, it is possible to further elaborate on one specific criterion, e.g., *treatability*, to determine the eligibility of disorders for NBS. Our study shows that it is very valuable to have such a discussion on *treatability* in the most transparent way, but also shows that this cannot be achieved without also addressing the other criteria in similar processes. Most criteria are interrelated; therefore, we consider this study as a starting point to help select disorders for NBS in an era in which NGS techniques increase the number of IMDs and other genetic diseases that are technically eligible for NBS. At the same time, we are on the brink of an era in which more and more (genetic) disorders will become increasingly treatable. Since this is the first study, to our knowledge, ever investigating a more transparent process for eligibility for NBS, this approach needs to be fine-tuned. We envision a funnel procedure to evaluate the eligibility of a disease by passing each of the W&J criteria and/or Andermann one by one, following the same procedure for every criterion and every disease.

Author Contributions: Conceptualization A.V., F.J.v.S. and M.R.H.-F.; methodology; A.V., F.J.v.S. and M.R.H.-F.; writing—original draft preparation, A.V.; writing—review and editing, F.J.v.S. and M.R.H.-F.; visualization, A.V.; supervision, F.J.v.S. and M.R.H.-F.; Members of the expert panel M.B.G.K., T.G.J.D., C.M.A.L., A.B.S., A.M.B., F.A.W., M.F.M., M.M.G.B., L.A.J.K., R.A.W., D.W., R.I.F.M.d.C., M.G.d.S.-v.d.V., K.K., S.A.F., G.V., J.M.P.v.d.H., H.H.H., J.S., B.P. and M.E.R.-G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding. None of the authors received financial funding for this Delphi study apart from the ZonMw grant (Grant No. 50-54300-98-506) received for the NGS4NBS project that aimed to identify and overcome technical challenges for genetic aspects of an NGS-based NBS program.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We would like to thank medical ethicists W. J. (Wybo) Dondorp and E. L. M. (Els) Maeckelberghe for their insights in the first draft of the statements.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A First Draft of the Statements by the Research Team and Medical Ethicists

Q1. To include a disorder in the NBS, there should be a treatment available that is:

- (a) Approved by the EMA
- (b) Covered by insurance companies (as set by the Dutch government)

Q2. What level of burden or risk for the newborn/child related to the available treatment is acceptable to include a disorder in NBS?

- (a) Little or no impact or burden with yearly blood tests, noninvasive imaging screening tests, oral medication with a low side-effect profile
- (b) Modest impact or burden with invasive screening tests, daily lifestyle/diet modification, and medication with a substantial side-effect profile
- (c) Significant impact or burden including intravenous treatment, removal, or transplantation of a non-vital organ (system) with frequent complications
- (d) Extreme impact or burden with transplantation of a vital organ (system)

Q3. To accept the inclusion of a disorder in the NBS, it is acceptable if a treatment works relatively well but still carries a risk for disease-related mortality

Q4. To accept the inclusion of a disorder in the NBS, it is acceptable if a treatment works relatively well but still carries a risk for disease-related morbidity

Q5. To accept the inclusion of a disorder in the NBS, early detection with consequent treatment, when compared to clinical presentation with consequent treatment, should at least:

- (a) Prevent (sudden) death
- (b) Prevent complications of all possible involved organs
- (c) Reverse all manifestations of the disorder
- (d) Prevent symptoms clearly related to one or two primary organ(s), while it is acceptable that treatment cannot prevent all clinical problems
- (e) Prevent developmental delay
- (f) Prevent recurrent admissions to the hospital
- (g) Improve the quality of life
- (h) Improve the quality-adjusted life-years (QALYs)

Q6. In how many publications should this outcome have been demonstrated, to accept the inclusion of a disorder in the NBS?

- (a) 1 Publication of 1 institute
- (b) At least 2 publications of 1 institute
- (c) 1 Publication of at least 2 institutes each
- (d) At least 2 Publication of at least 2 institutes each
- (e) At least. . .

Q7. To accept the inclusion of a disorder in the NBS, in what percentage of patients (both mild and severe variants) should (any) effect of treatment have been demonstrated?

- (a) In <25% of patients
- (b) In 25–50% of patients
- (c) In 50–75% of patients
- (d) In >75% of patients
- (e) At least in. . .

Appendix B Treatability Statements after Part 1.1

Q1. To include a disorder in the NBS, there should be a treatment available that is financially covered or reimbursed by standard health care:

Q2. What level of burden or risk for the newborn/child related to the available treatment is acceptable to include a disorder in NBS?

- (a) As long as the expected benefit/burden ratio of early treatment is positive for patients, any available treatment enabling NBS is acceptable

- (b) The burden of treatment is not important in determining treatability for NBS
- (c) In the decision to include a disorder in NBS the risk of overtreatment is less important than the risk of undertreatment

Q3. To include a disorder in the NBS, it is acceptable if a treatment works relatively well but still carries a small risk for disease-related mortality

Q4. Combined with Q5

Q5. To accept the inclusion of a disorder in the NBS, early detection with consequent treatment, when compared to clinical presentation with consequent treatment, should at least:

- (a) Prevent (sudden) death
- (b) Prevent complications of all possible involved organs
- (c) Reverse all manifestations of the disorder
- (d) Prevent symptoms clearly related to one or two primary organ(s), while it is acceptable that treatment cannot prevent all clinical problems
- (e) Prevent developmental delay
- (f) Prevent recurrent admissions to the hospital
- (g) Improve the quality of life
- (h) Improve the quality-adjusted life-years (QALYs)

Q6. In how many publications should this outcome have been demonstrated to accept the inclusion of a disorder in the NBS?

- (a) In very rare disorders at least one patient should be treated with any measurable effect with quality data about a plausible mechanism and/or animal models.
- (b) At least one paper on a new treatment by one institute with good quality data about a plausible mechanism with an “adequate number of patients” with clear effect size and suggesting the efficacy of early treatment is sufficient to accept that treatment for NBS.
- (c) Papers on a new treatment by at least two institutes with good quality data about a plausible mechanism with an “adequate number of patients” with clear effect size and suggesting the efficacy of early treatment is sufficient to accept that treatment for NBS.
- (d) Papers on a new treatment by at least three institutes with good quality data about a plausible mechanism with an “adequate number of patients” with clear effect size and suggesting the efficacy of early treatment is sufficient to accept that treatment for NBS.
- (e) There must be consensus on the positive outcome by an (international) expert meeting

Q7. In what percentage of patients (both mild and severe variants) should the expected effect of treatment of a specific disorder have been demonstrated to include a disorder in the NBS?

- (a) In <50% of patients
- (b) In >50% of patients
- (c) In >75% of patients
- (d) Basically all patients

Appendix C Comments from the Expert Panel on Part 2

IMD 1—PDE in NBS

“Proven treatment efficacy. Readily available treatment”.

“In PDE you want to treat seizures but also PMR, for the last there is evidence but may be less clear and spectacular than for seizure control. It would help to answer these questions if the latest papers are provided (if you do not know the last up-to-date info for this disease)”.

“We test consistently for PDE (and treat accordingly) in infants with unexplained epilepsy, not sure if implementation in NBS would lead to earlier detection/treatment”.

“Not financially covered or reimbursed for adults (yet) and almost no one will be capable of living a normal independent life. Mortality prevention is not reason enough to include a disorder in NBS. The treatment of epilepsy could also be done by changing guidelines, that make sure that every child with epilepsy will be screened directly for this disorder”.

“I missed questions on the appropriateness of biomarkers in body fluids”.

“No doubt”.

“Yet unknown, insufficient data with a possible positive response but in only nine patients, future studies are necessary as well as a valid screening parameter”.

“Yes, but is proven effective screening is still lacking”!

IMD 2—CG in NBS

“Relevance of NBS for CG is undisputed in my opinion”.

“Questionable what ‘a patient’ is, Duarte variants included or not”?

“I actually do not know if there is a treatment effect in > 75% of patients, I assume there is but I don’t know if this is studied that way”.

“Although the appearance of long-term complications are not affected by current intervention after NBS it can be lifesaving for the child with neonatal illness to know that it is due to CG because NBS is positive (avoids the chance that this diagnosis is overlooked with fatal consequences)”.

“The issue is that death is prevented and liver disease as well but mental retardation is not influenced really”.

“No doubt”.

“For reasons of the life-threatening newborn crisis, it does not prevent developmental delay and other long-term effects”.

“A big problem is that early disease and possible death, as well as cataract, are prevented BUT the treatment does not prevent significant developmental delay in many patients”.

IMD 3—CPT2 in NBS

“More studies needed to confirm NBS efficacy on lowering disease-related mortality, but in principle a good candidate for NBS”.

“Exceedingly rare, but treatable, should be included don’t know about how many patients have been treated, and how many mild patients”.

“I don’t know how many centers have thought about this treatment or whether this builds on knowledge for other fatty acid oxidation disorders”.

“Yes, if one can differentiate between myopathic and the other forms”.

“One of the issues will be that not all CPT2 patients need treatment from birth onwards and that you will treat persons unnecessarily with a mild disease as patients”.

“Research is necessary to investigate this. There is a discrepancy between severe phenotypes and late presentation. More research is needed to look e.g., how to discriminate between the severity of phenotypes”.

“Most cases are milder cases”.

“Easily treatable, for the ones at risk for hypoglycemia”.

“More study is needed to make sure if CPT2 should be added to NBS. This will at least consist of a cohort study including all Dutch patients, and if numbers are too low, an international study”.

IMD 4—GSDII in NBS

“Starting shortly after birth in classic infantile-onset Pompe disease will positively affect treatment outcome. Effect of NBS on development of long term complications likely limited”.

“Don’t know about consensus meetings. don’t know about diagnostic assays”.

“I think early treatment helps but does not prevent all disease manifestations. I think it might also help in the diagnostic delay and uncertainty in parents”.

“Questions about inclusion will not be related to treatability”.

“No because we cannot filter out the late-onset forms in an efficient way in a neonatal blood spot; not even with DNA techniques”.

“If detection of mild (late or adult-onset) variants can be prevented”.

IMD 5—GCH1 in NBS

“Detection of phenylalanine in NBS unsuitable test as AR form can be missed. Most symptoms can be reversed with treatment after diagnosis, even after a large diagnostic delay”.

“Don’t know about variants, don’t know about literature”.

“Also for this disease, I don’t have all the UpToDate knowledge but I guess there has not been an international consensus meeting”.

“Based on treatability this could fly in but we have to accept that MR cannot be easily prevented as in PKU”.

“I have no experience with this disorder. At large risk at over-diagnosing in my opinion”.

“For me, that would depend on the number of patients with biallelic mutations versus cases with dominant inheritance. That complicates the decision for defects in this gene. For this disease, I do have not enough overview of the current literature to decide on this question”.

“Insufficient data yet”.

“Insufficient knowledge about this disorder”.

IMD 6—OTC in NBS

“Patients still exhibit many metabolic decompensations despite early detection. If other treatment modalities become standard of care (liver Tx, gene therapy), the efficacy of NBS for OTC deficiency may become more apparent”.

“I am afraid that the most severe type will already give symptoms before the neonatal screening result and therefore you won’t fully prevent death or developmental delay, also diet and scavengers will not fully prevent this”.

“Not taken into account that it is X-linked affected girls: risk/benefit longitudinal data needed the question will be how you aim to tackle the difference between the very severe presentation of male newborns with usually later presenting female children”.

“For severe cases, NBS results may come too late”.

“There are insufficient good treatment options but harm can definitely be avoided and parents could be counseled for subsequent children. But that is not the aim of NBS”.

“In general X-linked diseases are more difficult for NBS. I see too many drawbacks”.

“Not the severe OTC in boys milder forms may benefit”.

“As we are discussing only the severe end of the phenotypic spectrum, I think screening will be too late to prevent the first clinical metabolic decompensation”.
 “It is the first decompensation that often determines the outcome of cognitive and motor impairment”.

IMD 7—WD in NBS

“Early detection and treatment will prevent complications of WD, but optimization of treatment still needed”.

“I am insufficiently aware of this condition and screening & treatment to comment”.

“Which patients are the true patients in need of treatment? If that can be answered then NBS is justified”.

“Looking back to Pompe disease, I guess there was an extra question on Qualys not included for other diseases, so should be left out in Pompe”.

“I lack experience in the various forms (liver/neurologic) of Wilson’s disease and effects of therapy”.

“No doubt”.

IMD 8—MCM in NBS

“Still significant disease burden despite early detection, the effect of long term complications (i.e., kidney failure) unclear”.

“Already part of NBS”.

“The available treatment helps but does not fully prevent longterm complications in MMA in addition when the NBS results are available the most severe MMA patients are often already symptomatic so yes I think NBS helps but does not prevent all MMA complications”.

“We do not -by the method now at the time performed now- achieve that we will prevent death etc. in severe forms, but if it will be early enough it should prevent much of the disease. So, the answer depends on the interpretation of the question”.

“Unfortunately, the effects of early treatment (when comparing early identified patients by screening with their clinically identified sib) are most effective in preventing the initial presentation and consequences thereof”.

“I am not fully convinced”.

“Not for the severe MMA mut 0 or mut—as they will have presented already the milder CBLA CBLB will have a major benefit”.

“Severe MCM is not fit for inclusion in NBS. However, the milder causes of MMA are”!

IMD 9—TH in NBS

“Currently unclear how NBS for TH would affect the clinical outcome. Variable response to treatment”.

“Yes, but not all types will have full treatment effect (for the severe type this effect can be limited) in addition I don’t know all the evidence for this disease”.

“Limited experience”.

“Difficult to diagnose, a long delay and treatment has a very good effect, that is why I have chosen this”

“No doubt”.

IMD 10—PKU in NBS

“Proven benefit. No discussion on NBS”.

“Included as positive control? I indeed scored high in all aspects of treatability”

“No doubt”.

Appendix D Comments from the Expert Panel on the Final Treatability Statements

Opinion 1: It can be a useful tool if all data is available and experts are involved in decision-making.

“Yes I think so but to fill in you have to know the up-to-date knowledge and I think you can better fill this in if you have read the relevant literature beforehand”.

“Yes, but must be performed by (international) experts in treatment of that particular disease”.

“It can be”.

“Yes if all the data (evidence for consensus meetings etc.) is easily available”.

“Overall this is a difficult subject. On a few of these diseases, I am not an expert, and in my opinion, experts should describe the effect of early treatment and then together with these experts we can make a decision”.

Opinion 2: It can be useful tool as a starting point for a discussion for inclusion in NBS.

“It would be a useful tool to start the discussion”.

“I guess so, counting the produced marks together. Would think it can be used as a pyramid. If you have gone through this level the matrix using the other Wilson & Jungner criteria can be put in a matrix receiving marks/points to pass the criteria to be included”.

Opinion 3: Concerns about how to proceed with this tool if disorders are not eligible for NBS.

“Yes, this may work. I had some troubles with the ‘dubbele ontkenning’ (double denial?), which is when you decide an IMD is NOT fit for inclusion, and then you have to state how important ‘the presence of an effective treatment’ is, as for some there is no effective treatment, so you tend to say that it is not important, while it actually is (though there is not a treatment). Rather difficult to write this down, but I hope you get my point.

“In particular the questions are not suitable to help to support the advice for not selecting the disease for NBS”.

Opinion 4: Concerns about how to proceed with this tool in disorders with a broad phenotypic variability.

“In addition, it is kind of difficult not to think about severe vs. less severe or even adult-onset disease. If I am correct it was stated in the introduction that we should think about the severe phenotypes when doing this survey, but at the end of the matrix, it suddenly states ‘severe and less severe’ (or something like that). Finally, quality of life is a very complex issue here. How to measure this? How reliable? Disease-specific scales, etc., etc”.

“It seems useful for diseases with limited phenotypic variability. For diseases with a wider phenotypic variability, it seems more difficult because the two ends of the spectrum cannot be compared and usually not enough data are available on late-onset disease. Example: Pompe disease.”.

Opinion 5: The decision matrix is not useful (yet) to quantify treatability for IMDS in NBS

“No, too oversimplified, without profound knowledge of every disorder”.

“Sometimes questions are composed of more statements and therefore are difficult to answer clearly. Some questions are still difficult to understand, e.g., the question: ‘The treatment/diet results. . . disease-related mortality.’ This matrix is rather global and not all important issues per disease are addressed. In my opinion, the clarity/interpretability of this final decision-matrix is not yet sufficient to be useful to quantify treatability for IMDs in NBS”.

“In general, yes, but:—not all questions applicable to all diseases (e.g., preventing (acute) death, developmental delay). ‘Not applicable’ could be added as an option.—due to limited knowledge on some diseases several questions hard to answer.—question should be added about the level of expertise of the respondent with the specific disorder which is assayed”.

References

1. Wilson, J.M.G.; Jungner, G.; World Health Organization. *Principles and Practice of Screening for Disease*; World Health Organization: Geneva, Switzerland, 1968.
2. Andermann, A.; Blancquaert, I.; Beauchamp, S.; Déry, V. Revisiting Wilson and Jungner in the genomic age: A review of screening criteria over the past 40 years. *Bull. World Health Organ.* **2008**, *86*, 317–319. [[CrossRef](#)]
3. Andermann, A.; Blancquaert, I.; Beauchamp, S.; Costea, I. Guiding policy decisions for genetic screening: Developing a systematic and transparent approach. *Public Health Genom.* **2011**, *14*, 9–16. [[CrossRef](#)] [[PubMed](#)]
4. Sturdy, S.; Miller, F.; Hogarth, S.; Armstrong, N.; Chakraborty, P.; Cressman, C.; Dobrow, M.; Flitcroft, K.; Grossman, D.; Harris, R.; et al. Half a Century of Wilson & Jungner: Reflections on the Governance of Population Screening. *Wellcome Open Res.* **2020**, *5*, 158. [[PubMed](#)]
5. Loeber, J.G.; Platis, D.; Zetterström, R.H.; Almashanu, S.; Boemer, F.; Bonham, J.R.; Borde, P.; Brincat, I.; Cheillan, D.; Dekkers, E.; et al. Neonatal Screening in Europe Revisited: An ISNS Perspective on the Current State and Developments Since 2010. *Int. J. Neonatal Screen.* **2021**, *7*, 15. [[CrossRef](#)] [[PubMed](#)]
6. Watson, M.S.; Mann, M.Y.; Lloyd-Puryear, M.A.; Rinaldo, P.; Howell, R.R. Newborn Screening: Toward a Uniform Screening Panel and System—Executive Summary. *Pediatrics* **2006**, *117* (Suppl. S3), S296–S307. [[CrossRef](#)]
7. Kemper, A.R.; Green, N.S.; Calonge, N.; Lam, W.K.; Comeau, A.M.; Goldenberg, A.J.; Ojodu, J.; Prosser, L.A.; Tanksley, S.; Bocchini, J.A., Jr. Decision-making process for conditions nominated to the Recommended Uniform Screening Panel: Statement of the US Department of Health and Human Services Secretary’s Advisory Committee on Heritable Disorders in Newborns and Children. *Genet. Med.* **2014**, *16*, 183–187. [[CrossRef](#)]
8. Children ACoHDiNa. *Recommended Uniform Screening Panel*; Health Resources & Services Administration: Rockville, MD, USA, 2018.
9. Lund, A.M.; Wibrand, F.; Skogstrand, K.; Bækvad-Hansen, M.; Gregersen, N.; Andresen, B.S.; Hougaard, D.M.; Dunø, M.; Olsen, R.K.J. Use of Molecular Genetic Analyses in Danish Routine Newborn Screening. *Int. J. Neonatal Screen.* **2021**, *7*, 50. [[CrossRef](#)]
10. Australian Health Ministers’ Advisory Council. *Newborn Bloodspot Screening National Policy Framework*; Australian Health Ministers’ Advisory Council: Canberra, Australia, 2018.
11. Committee CP; Screening SCO. *Population-Based Screening Framework*; Australian Government Department of Health and Aged Care: Canberra, Australia, 2018.
12. Australian Government. *NBS Conditions Screened in Australia, September 2022*; Australian Government Department of Health and Aged Care: Canberra, Australia, 2022.
13. Pollitt, R.J. Introducing new screens: Why are we all doing different things? *J. Inherit. Metab. Dis.* **2007**, *30*, 423–429. [[CrossRef](#)]
14. Furnier, S.M.; Durkin, M.S.; Baker, M.W. Translating Molecular Technologies into Routine Newborn Screening Practice. *Int. J. Neonatal Screen.* **2020**, *6*, 80. [[CrossRef](#)]
15. Adhikari, A.N.; Gallagher, R.C.; Wang, Y.; Currier, R.J.; Amatuni, G.; Bassaganyas, L.; Chen, F.; Kundu, K.; Kvale, M.; Mooney, S.D.; et al. The role of exome sequencing in newborn screening for inborn errors of metabolism. *Nat. Med.* **2020**, *26*, 1392–1397. [[CrossRef](#)]
16. Holm, I.A.; Agrawal, P.B.; Ceyhan-Birsoy, O.; Christensen, K.D.; Fayer, S.; Frankel, L.A.; Genetti, C.A.; Krier, J.B.; LaMay, R.C.; Levy, H.L.; et al. The BabySeq project: Implementing genomic sequencing in newborns. *BMC Pediatr.* **2018**, *18*, 225. [[CrossRef](#)] [[PubMed](#)]
17. Milko, L.V.; O’Daniel, J.M.; DeCristo, D.M.; Crowley, S.B.; Foreman, A.K.M.; Wallace, K.E.; Mollison, L.F.; Strande, N.T.; Ginary, Z.S.; Boshe, L.J.; et al. An Age-Based Framework for Evaluating Genome-Scale Sequencing Results in Newborn Screening. *J. Pediatr.* **2019**, *209*, 68–76. [[CrossRef](#)] [[PubMed](#)]

18. Veldman, A.; Kiewiet, M.B.G.; Heiner-Fokkema, M.R.; Nelen, M.R.; Sinke, R.J.; Sikkema-Raddatz, B.; Voorhoeve, E.; Westra, D.; Dollé, M.E.T.; Schielen, P.C.J.I.; et al. Towards Next-Generation Sequencing (NGS)-Based Newborn Screening: A Technical Study to Prepare for the Challenges Ahead. *Int. J. Neonatal Screen.* **2022**, *8*, 17. [[CrossRef](#)] [[PubMed](#)]
19. Kingsmore, S.F.; Smith, L.D.; Kunard, C.M.; Bainbridge, M.; Batalov, S.; Benson, W.; Blicow, E.; Caylor, S.; Chambers, C.; Del Angel, G.; et al. A genome sequencing system for universal newborn screening, diagnosis, and precision medicine for severe genetic diseases. *Am. J. Hum. Genet.* **2022**, *109*, 1605. [[CrossRef](#)] [[PubMed](#)]
20. Van Karnebeek, C.D.M.; Shevell, M.; Zschocke, J.; Moeschler, J.B.; Stockler, S. The metabolic evaluation of the child with an intellectual developmental disorder: Diagnostic algorithm for identification of treatable causes and new digital resource. *Mol. Genet. Metab.* **2014**, *111*, 428–438. [[CrossRef](#)]
21. Van Konijnenburg, E.M.M.H.; Wortmann, S.B.; Koelewijn, M.J.; Tseng, L.A.; Houben, R.; Stöckler-Ipsiroglu, S.; Ferreira, C.R.; van Karnebeek, C.D.M. Treatable inherited metabolic disorders causing intellectual disability: 2021 review and digital app. *Orphanet J. Rare Dis.* **2021**, *16*, 170. [[CrossRef](#)]
22. Jünger, S.; Payne, S.A.; Brine, J.; Radbruch, L.; Brearley, S.G. Guidance on Conducting and REporting DELphi Studies (CREDES) in palliative care: Recommendations based on a methodological systematic review. *Palliat. Med.* **2017**, *31*, 684–706. [[CrossRef](#)]
23. Veugelers, R.; Gaakeer, M.I.; Patka, P.; Huijsman, R. Improving design choices in Delphi studies in medicine: The case of an exemplary physician multi-round panel study with 100% response. *BMC Med. Res. Methodol.* **2020**, *20*, 156. [[CrossRef](#)]
24. Wainwright, P.; Gallagher, A.; Tompsett, H.; Atkins, C. The use of vignettes within a Delphi exercise: A useful approach in empirical ethics? *J. Med. Ethics* **2010**, *36*, 656–660. [[CrossRef](#)]
25. Vogel, C.; Zwolinsky, S.; Griffiths, C.; Hobbs, M.; Henderson, E.; Wilkins, E. A Delphi study to build consensus on the definition and use of big data in obesity research. *Int. J. Obes.* **2019**, *43*, 2573–2586. [[CrossRef](#)]
26. Beiderbeck, D.; Frevel, N.; Von Der Gracht, H.A.; Schmidt, S.L.; Schweitzer, V.M. Preparing, conducting, and analyzing Delphi surveys: Cross-disciplinary practices, new directions, and advancements. *MethodsX* **2021**, *8*, 101401. [[CrossRef](#)] [[PubMed](#)]
27. Keeney, S.; Hasson, F.; McKenna, H.P. *The Delphi Technique in Nursing and Health Research*; Wiley-Blackwell: Chichester, UK, 2011.
28. Hasson, F.; Keeney, S.; McKenna, H. Research guidelines for the Delphi survey technique. *J. Adv. Nurs.* **2000**, *32*, 1008–1015. [[CrossRef](#)] [[PubMed](#)]
29. Tomasik, T. Reliability and validity of the Delphi method in guideline development for family physicians. *Qual. Prim. Care* **2010**, *18*, 317–326. [[PubMed](#)]
30. Ceyhan-Birsoy, O.; Murry, J.B.; Machini, K.; Lebo, M.S.; Yu, T.W.; Fayer, S.; Genetti, C.A.; Schwartz, T.S.; Agrawal, P.B.; Parad, R.B.; 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. [[CrossRef](#)]
31. Wojcik, M.H.; Zhang, T.; Ceyhan-Birsoy, O.; Genetti, C.A.; Lebo, M.S.; Yu, T.W.; Parad, R.B.; Holm, I.A.; Rehm, H.L.; Beggs, A.H.; et al. Discordant results between conventional newborn screening and genomic sequencing in the BabySeq Project. *Genet. Med.* **2021**, *23*, 1372–1375. [[CrossRef](#)]
32. Keeney, S.; Hasson, F.; McKenna, H. Consulting the oracle: Ten lessons from using the Delphi technique in nursing research. *J. Adv. Nurs.* **2006**, *53*, 205–212. [[CrossRef](#)]
33. Niederberger, M.; Spranger, J. Delphi Technique in Health Sciences: A Map. *Front. Public Health* **2020**, *8*, 457. [[CrossRef](#)]
34. Hohmann, E.; Brand, J.C.; Rossi, M.J.; Lubowitz, J.H. Expert Opinion Is Necessary: Delphi Panel Methodology Facilitates a Scientific Approach to Consensus. *Arthrosc. J. Arthrosc. Relat. Surg.* **2018**, *34*, 349–351. [[CrossRef](#)]
35. Harbour, R.; Miller, J. A new system for grading recommendations in evidence based guidelines. *BMJ Br. Med. J.* **2001**, *323*, 334. [[CrossRef](#)]
36. Guyatt, G.H.; Oxman, A.D.; Vist, G.E.; Kunz, R.; Falck-Ytter, Y.; Alonso-Coello, P.; Schünemann, H.J. GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* **2008**, *336*, 924. [[CrossRef](#)]
37. Boulkedid, R.; Abdoul, H.; Loustau, M.; Sibony, O.; Alberti, C. Using and Reporting the Delphi Method for Selecting Healthcare Quality Indicators: A Systematic Review. *PLoS ONE* **2011**, *6*, e20476. [[CrossRef](#)] [[PubMed](#)]
38. Markmann, C.; Spickermann, A.; Von Der Gracht, H.A.; Brem, A. Improving the question formulation in Delphi-like surveys: Analysis of the effects of abstract language and amount of information on response behavior. *Futures Foresight Sci.* **2021**, *3*, 56. [[CrossRef](#)]
39. Yaniv, I. Group diversity and decision quality: Amplification and attenuation of the framing effect. *Int. J. Forecast.* **2011**, *27*, 41–49. [[CrossRef](#)]
40. Winkler, J.; Moser, R. Biases in future-oriented Delphi studies: A cognitive perspective. *Technol. Forecast. Soc. Change* **2016**, *105*, 63–76. [[CrossRef](#)]
41. Van Dijk, T.; Kater, A.; Jansen, M.; Dondorp, W.J.; Blom, M.; Kemp, S.; Langeveld, M.; Cornel, M.C.; van der Pal, S.M.; Henneman, L. Expanding Neonatal Bloodspot Screening: A Multi-Stakeholder Perspective. *Front. Pediatr.* **2021**, *9*, 706394. [[CrossRef](#)] [[PubMed](#)]

42. Armstrong, B.; Christensen, K.D.; Genetti, C.A.; Parad, R.B.; Robinson, J.O.; Zawatsky, C.L.B.; Zettler, B.; Beggs, A.H.; Holm, I.A.; Green, R.C.; et al. Parental Attitudes Toward Standard Newborn Screening and Newborn Genomic Sequencing: Findings from the BabySeq Study. *Front. Genet.* **2022**, *13*, 867371. [[CrossRef](#)] [[PubMed](#)]
43. Blom, M.; Bredius, R.G.M.; Jansen, M.E.; Weijman, G.; Kemper, E.A.; Vermont, C.L.; Hollink, I.H.I.M.; Dik, W.A.; van Montfrans, J.M.; van Gijn, M.E.; et al. Parents' Perspectives and Societal Acceptance of Implementation of Newborn Screening for SCID in the Netherlands. *J. Clin. Immunol.* **2021**, *41*, 99–108. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.