Title

Navigating the outcome maze: a scoping review of outcomes and instruments in clinical trials in genetic neurodevelopmental disorders and intellectual disability.

Authors

Müller, Annelieke

van Silfhout, Nadia

den Hollander, Bibiche

et al.

Publication Date

2024

DOI

10.1177/26330040241245721

Peer reviewed

THERAPEUTIC ADVANCES in
Rare Disease

Original Research

Navigating the outcome maze: a scoping review of outcomes and instruments in clinical trials in genetic neurodevelopmental disorders and intellectual disability

Ther Adv Rare Dis

2024, Vol. 5: 1–19

DOI: 10.1177/26330040241245721

Annelieke R. Müller, Nadia Y. van Silfhout, Bibiche den Hollander, Dick H. C. Kampman, Lianne Bakkum, Marion M. M. G. Brands, Lotte Haverman, Caroline B. Terwee, Carlo Schuengel, Joost Daams, David Hessl, Frits A. Wijburg, Erik Boot and Agnies M. van Eeghen

Abstract

Background: Individuals with genetic neurodevelopmental disorders (GNDs) or intellectual disability (ID) are often affected by complex neuropsychiatric comorbidities. Targeted treatments are increasingly available, but due to the heterogeneity of these patient populations, choosing a key outcome and corresponding outcome measurement instrument remains challenging.

Objectives: The aim of this scoping review was to describe the research on outcomes and instruments used in clinical trials in GNDs and ID.

Eligibility criteria: Clinical trials in individuals with GNDs and ID for any intervention over the past 10 years were included in the review.

Sources of evidence: MEDLINE, PsycINFO, and Cochrane CENTRAL were searched. Titles and abstracts were independently screened for eligibility with a subsample of 10% double-screening for interrater reliability. Data from full texts were independently reviewed. Discrepancies were discussed until consensus was reached.

Charting methods: Information was recorded on patient populations, interventions, designs, outcomes, measurement instruments, and type of reporter when applicable. Qualitative and descriptive analyses were performed.

Results: We included 312 studies reporting 91 different outcomes, with cognitive function most frequently measured (28%). Various outcome measurement instruments ($n = 457$) were used, with 288 in only a single clinical trial. There were 18 genetic condition-specific instruments and 16 measures were designed ad-hoc for one particular trial. Types of report included proxy-report (39%), self-report (22%), clinician-report (16%), observer-report (6%), self-assisted report (1%), or unknown (16%).

Conclusion: This scoping review of current practice reveals a myriad of outcomes and outcome measurement instruments for clinical trials in GNDs and ID. This complicates generalization, evidence synthesis, and evaluation. It underlines the need for consensus on suitability, validity, and relevancy of instruments, ultimately resulting in a core outcome set. A series of steps is proposed to move from the myriad of measures to a more unified approach.

Correspondence to:
Agnies M. van Eeghen
Emma Center for Personalized Medicine, Amsterdam UMC, Meibergdreef 9, Amsterdam, 1105 AZ, The Netherlands

Advisium, ´s Heeren Loo, Berkenweg 11, 3818 LA, Amersfoort, The Netherlands

Amsterdam Public Health Research Institute, Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands

Amsterdam Reproduction & Development, Child Development, Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands
a.m.vaneeghen@amsterdamumc.nl

Annelieke R. Müller
Department of Pediatrics, Emma Children´s Hospital, Amsterdam Gastroenterology Endocrinology Metabolism, Amsterdam UMC Location University of Amsterdam, Amsterdam, The Netherlands

Advisium, ´s Heeren Loo, Amersfoort, Utrecht, The Netherlands

Amsterdam Public Health Research Institute, Amsterdam, The Netherlands

Emma Center for Personalized Medicine, Amsterdam UMC, Amsterdam, The Netherlands

Nadia Y. van Silfhout
Lotte Haverman
Amsterdam Public Health Research Institute, Amsterdam, The Netherlands

Amsterdam Reproduction & Development, Child Development, Amsterdam, The Netherlands

Emma Children's Hospital, Child and Adolescent Psychiatry & Psychosocial Care, Amsterdam UMC Location University of Amsterdam, Amsterdam, The Netherlands

Bibiche den Hollander
Department of Pediatrics, Emma Children's Hospital, Amsterdam Gastroenterology Endocrinology Metabolism, Amsterdam UMC Location University of Amsterdam, Amsterdam, The Netherlands

Emma Center for Personalized Medicine, Amsterdam UMC, Amsterdam, The Netherlands

United for Metabolic Diseases, Amsterdam, The Netherlands

Dick H. C. Kampman
Faculty of Science, Operational Management, ICT Department, Utrecht University, Utrecht, The Netherlands

Lianne Bakkum
Carlo Schuengel
Amsterdam Public Health Research Institute, Amsterdam, The Netherlands

Department of Clinical Child and Family Studies, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Marion M. M. G. Brands
Department of Pediatrics, Emma Children's Hospital, Amsterdam Gastroenterology Endocrinology Metabolism, Amsterdam UMC Location University of Amsterdam, Amsterdam, The Netherlands

Emma Center for Personalized Medicine, Amsterdam UMC, Amsterdam, The Netherlands

Plain Language Summary

Navigating the maze of outcome measures in rare disorders

Treatments for genetic neurodevelopmental disorders and intellectual disability are increasingly available. However, it is hard to find appropriate instruments to measure whether these treatments are working. This hampers research and means some patients might not get the treatment they need. This scoping review provides an overview of investigated outcomes in this group, and with which instruments these are measured. It reveals that many different and overlapping outcomes are measured, complicating gathering, combining, and comparing of evidence. This scoping review underlines the need for harmonization and consensus on suitability, validity, and relevancy. Steps are proposed to move from the maze of outcome measures to a unified approach. Also, we provided recommendations for researchers to measure what matters to affected individuals and patient-centered care.

Keywords: intellectual disability, measures, outcomes, psychiatry, quality of life, rare genetic disorders

Received: 18 October 2023; revised manuscript accepted: 14 March 2024.

Introduction

Intellectual disability (ID) occurs in 1–3% of the population and is characterized by substantial limitations in both intellectual functioning and adaptive behavior, originating during the developmental period.[1–3] Exogenous factors such as an infection and birth complications may cause ID,[4] and with novel techniques such as exome and genome sequencing, a genetic etiology can be identified in up to 50% of the individuals with ID with many more awaiting diagnosis[5,6] (Figure 1). Although these genetic neurodevelopmental disorders (GNDs), including syndromic ID and neurometabolic disorders, are individually rare, collectively they are common.[7,8] In GNDs, the level of intellectual functioning is variable, ranging from normal or borderline functioning to profound ID.[9–11] Although GNDs and ID

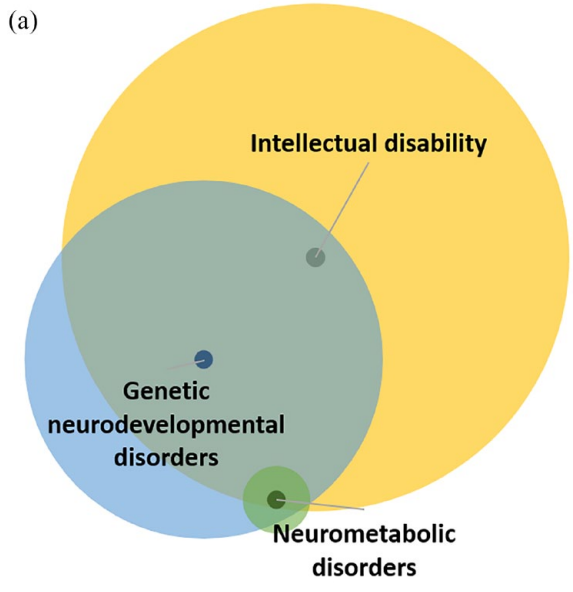

Figure 1. *(Continued)*

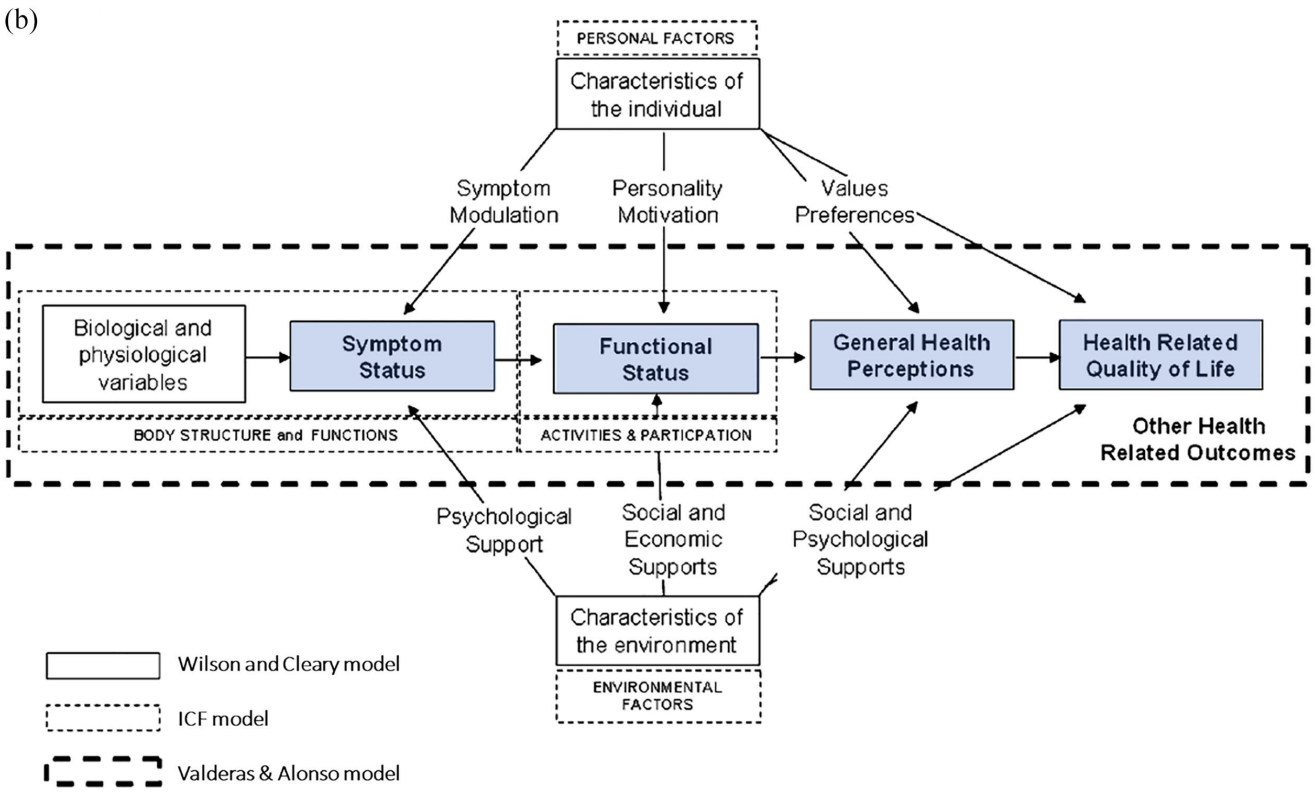

Figure 1. (a) Schematic representation of target populations. Importantly, it represents an indication rather than a precise scaled proportion of target populations. (b) Domains of interest (blue boxes) to this review (symptom and functional status related to neurological functioning, and the overarching concepts of perceived health and overall quality of life). This is based on the conceptual model of health outcomes from Valderas and Alonso which incorporates both the commonly used models of the ICF and Wilson and Cleary.[12,13] Figure adapted from Valderas and Alonso.
ICF, International Classification of Functioning, disability and health.

populations have often been separately studied, there is substantial overlap in patient populations.

Individuals with GND and ID are often affected by complex somatic and neuropsychiatric comorbidity, with great inter- and intra-individual variability. Neuropsychiatric manifestations typically cause the greatest burden for the affected individual, their families, and on healthcare systems, with a substantial clinical and economic burden.[14] The Food and Drug Administration (FDA) considers clinical outcome assessments (COAs), including patient-reported outcomes (PROs), clinician-reported outcomes (ClinROs), observer-reported outcomes (ObsROs), and performance outcomes (PerfOs), well-defined and reliable assessments of affected individuals' symptoms, overall mental state, or how they function (Box 1).[15–17]

Knowledge about the genetic etiology of GNDs rapidly increases and offers disorder-specific treatment options which can be targeted to the

Amsterdam Reproduction & Development, Child Development, Amsterdam, The Netherlands

United for Metabolic Diseases, Amsterdam, The Netherlands

Caroline B. Terwee
Amsterdam Public Health Research Institute, Amsterdam, The Netherlands

Box 1. Definitions and abbreviations of commonly used terminology with regard to outcomes and outcome measurement instruments.[12,15,18,19]

Definition	Abbreviation	Explanation
Clinical outcome assessment*	COA	A clinical outcome assessment describes or reflects how a person feels, functions, or survives and can be reported by the affected individual, a non-clinical observer (such as parent), a health-care provider, or through performance of an activity or task.
Outcome		An outcome refers to a construct or domain. In the context of a clinical trial, it refers to what is being measured on trial participants to examine the effect of exposure to a health intervention (e.g. anxiety).
Outcome measurement instrument		An outcome measurement instrument specifically refers to how the outcome is being measured. It is a tool to measure a quality or quantity of the outcome. It can be used to identify meaningful change for the individual, evaluate the effect of interventions, demonstrate the impact and value of interventions, identify areas for improvement, and benchmark against other interventions. Power calculations are often based on the chosen primary outcome measure. In literature, the term outcome measure has often been inconsistently and interchangeably used to refer to both the outcome and outcome measurement instrument; we consider using 'outcome measure' as an abbreviation of 'outcome measure instrument'.
Patient-reported outcome	PRO	A type of clinical outcome assessment, based on a report that comes directly from the affected individual about the status of the health condition.
Patient-reported outcome measure*	PROM	Instrument or tool utilized to measure PROs to evaluate the affected individuals' health status from their perspective. For individuals with an ID who are not able to complete a measure, a PROM can also be a proxy-report provided that it is someone who knows the affected individual well and fills out the PROM from the affected individual's perspective.
Clinician-reported outcome*	ClinRO	A type of clinical outcome assessment, based on a report that comes from a trained health-care professional after observation of a patient's health condition.
Performance outcome*	PerfO	A type of clinical outcome assessment, based on standardized task(s) actively undertaken by an affected individual according to instructions.
Observer-reported outcome*	ObsRO	A type of clinical outcome assessment, based on a report of observable signs, events, or behaviors related to an affected individual's health condition by someone other than the affected individual or a health-care professional, such as a parent, teacher, or caregiver.
Proxy		Someone who reports an outcome as if they were the affected individual themselves. Proxies report on behalf of the affected individual, in contrast to an observer-report in which the informant provides information about the manifestations and condition.
Generic outcome measure		A measure for a health concept that is relevant to a wide range of patient groups, enabling aggregation and comparison across varied conditions and settings.
Condition-specific outcome measures		A measure capturing elements of health relevant to a particular patient group or designed for a specific patient population.
Personalized outcome measure		A measure that refers to an instrument in which the domains and/or weights are not fixed. Outcome areas are specific for each individual and the affected individual (or proxy) is involved in identifying and setting specific outcome areas. In clinical trials, these are intended for standardized evaluation of an intervention's effectiveness based on individualized problems or goals.

*Adapted from the FDA.

gene, protein, or downstream biological pathway.[20] It allows for personalized care, which is the implementation of etiology-drive health monitoring and treatments.[21] For neuropsychiatric manifestations, targeted treatments are underway[22–25] and guidelines are increasingly available.[26]

However, interventional research in GNDs and ID is challenging. This is due to the rarity, complexity, and variability of health manifestations, even among individuals with the same disorder, as well as the heterogeneity in treatment response.[27] Other hurdles in these populations include varying cognitive and adaptive abilities, environmental factors, high rate of behavioral and emotional disturbances, a lack of stability, practice effects, and lack of consensus on the best measures within a particular construct.[28] Many outcomes have been measured in the past, but assessments of disease severity using clinical rating scales omitted patient perspectives about issues of relevance to their health. Deciding upon an appropriate outcome measure can be a daunting task, taking into account the acceptability and feasibility, and important measurement properties, such as validity, reliability, and responsiveness to change.

Noticeably, selection of outcome measures for a study has far-reaching implications. Previous trials that did not demonstrate significant clinical benefits based on the primary endpoints have been deemed 'negative' or 'failed' even though improvement on secondary endpoints or in clinical subgroups may be present,[28] as happened for clinical trials investigating the effects of Arbaclofen in Fragile X syndrome.[29,30] As such, inappropriate outcomes or outcome measurement instruments can result in negative results about the effectiveness of interventions, potentially meaning that truly effective treatments do not become available to patients and their families.[20,31]

The aim of this scoping review was to provide an overview of outcomes and outcome measurement instruments selected in clinical trials in individuals with ID and GNDs, measured by COAs focusing on neurological functioning, mental and social functioning, and the overarching concepts of perceived health and health-related quality of life (HR-QoL). The findings from this scoping review may serve as a starting point for discussion about relevant outcomes and outcome measurement instruments and result in a series of steps and recommendations to move to a more unified approach in clinical trials in individuals with ID and GNDs.

Methods

We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) Protocols, and the PRISMA extension for Scoping Reviews (PRISMA-ScR) checklist[32,33] (see Supplemental Figure 1). The methodological framework was posted in advance on the Open Science Framework: https://osf.io/2zmxv/.

Eligibility

Clinical trials for any intervention were included in the review, including comparative studies (randomized and non-randomized), single-case trials and single-arm case series (retrospective and prospective), and trial protocols. Validation and feasibility studies, and economic evaluations were excluded. GNDs were defined as disorders with a genetic etiology affecting the nervous system in early development. GNDs associated with ID were included (Figure 1). ID was defined as substantial limitations in both intellectual functioning and adaptive behavior, originating during the developmental period.[1] Neurometabolic disorders, consisting of a subgroup of rare genetic hereditary conditions in which the impairment of a biochemical pathway is essential to the pathophysiology of the disease,[34] were included in case they are associated or presented with intellectual deficits/impairment. Studies were included when a participant showed intellectual impairments. Exclusion criteria included ID explicitly stated to be due to exogenous factors.

Studies were included if these used a COA, that is, a PRO, ClinRO, ObsRO, or PerfO, with regard to at least one domain of interest (Box 1; Figure 1). Condition-specific and personalized outcome measures were included as well (Box 1). Biological and physiological variables were excluded. Clinical trials with only epilepsy characteristics or motor function as outcome without another COA

Epidemiology and Data Science, Amsterdam UMC Location Vrije Universiteit, Amsterdam, The Netherlands

Joost Daams
Medical Library, Research Support, Amsterdam UMC Location University of Amsterdam, Amsterdam, The Netherlands

David Hessl
MIND Institute and Department of Psychiatry and Behavioral Sciences, University of California Davis, Sacramento, CA, USA

Frits A. Wijburg
Department of Pediatrics, Emma Children's Hospital, Amsterdam Gastroenterology Endocrinology Metabolism, Amsterdam UMC Location University of Amsterdam, Amsterdam, The Netherlands

Erik Boot
Advisium, 's Heeren Loo, Amersfoort, Utrecht, The Netherlands

The Dalglish Family 22q Clinic, Toronto, ON, Canada

Department of Psychiatry & Neuropsychology, Maastricht University Medical Center, Maastricht, The Netherlands

about mental or social functioning, general health perceptions, or HR-QoL were excluded to narrow the scope of the review. Eligible assessments included descriptions related to neurological functioning, mental and social functioning, and general health perceptions or HR-QoL.

Search strategy and study selection

MEDLINE (Ovid), PsycINFO, and Cochrane CENTRAL were searched from 2012 to 2022, with assistance of a clinical research librarian (JGD). A list of genetic disorders associated with ID was composed using the human phenotype ontology (HPO) database on https://hpo.jax.org/app/. All terms describing a genetic disease assigned to the subontology ID were included; collectively the HPO-ID list of GNDs. Furthermore, a search strategy for ID without known genetic etiology was used in combination with terms for trials (https://osf.io/2zmxv/). A time limit of the last 10 years was applied due to feasibility reasons and abundancy, and to identify the most recent clinical trials, as the field of trials for rare diseases is emerging quickly. Additional papers were identified by reference list checking. To enhance precision of search results, VOSviewer was used to visually identify potentially irrelevant terms eligible for exclusion with corresponding network visualization, contributing to disambiguation. These included Duchenne muscular dystrophy, retinitis pigmentosa, and Charcot-Marie-Tooth disease (Supplemental Material Figure 2). Prior to excluding a category, records were screened and checked to confirm exclusion criteria. When for a specific trial both a research article and trial protocol from a register were available, only the research article was included.

The application Rayyan was used for screening.[35] Titles and abstracts were independently screened for eligibility by six reviewers (AM, BdH, EB, LB, MB, and AvE) who all screened one-sixth of the selected items, with a subsample of 10% double-screening for interrater reliability. The proportion of discrepancies in the double-screening varied from 0% to 5%, mainly due to uncertainty about whether a specific disorder was associated with intellectual impairments. Discrepancies were discussed until consensus was reached. In case of uncertainty about whether the population met inclusion criteria, an expert with regard to the specific condition was consulted. Full texts were screened for eligibility, and data were independently reviewed by seven reviewers (AM, BdH, EB, LB, MB, NvS, and AvE) with a sample of 10% double-reviewing for interrater reliability. Potential discrepancies were solved through discussion.

Data extraction

The following data were extracted: title, year of publication, first author, journal, countries of study, type of study, number of participants, GND/neurometabolic disorder/heterogeneous ID of unknown cause (if neurometabolic disorder, this category was used; GNDs and neurometabolic disorders can overlap), diagnosis, patient characteristics (including age, sex, severity of ID), design, duration of trial, randomization, blinding, intervention, comparator used, type of COA, reported outcomes, outcome measurement instruments and version, whether it concerned a condition-specific or personalized outcome measure, type of report, number of assessments, mode of data collection, setting, and involvement of patient/parent perspectives regarding the choice of outcome measures. The reported outcomes were classified according to the most commonly used terms by the authors of the included studies. As for outcomes related to behavior, the term 'behavior' was used when general behavior was reported or when it was not further specified. Otherwise, terminology for specific behavior was reported, such as 'repetitive behavior'. When version of the outcome measurement instrument was specified, information is provided in the Supplemental tables when reported by the authors of the included clinical studies. Due to limited reporting and the large number of different instruments used, numbers corresponding to the outcome measurement instruments were used regardless of different versions.

Outcome measurement instruments were classified based on the reported outcomes and information provided in the articles using a conceptual model of health outcomes from Valderas and Alonso, which is a combination of the classification system of Wilson and Cleary and the International Classification of Functioning, Disability and Health [Figure 1(b)].[12,36] Domains included symptoms, physical function, mental function, social function, general health perceptions, and HR-QoL. Additionally, cognitive

function was included as a separate domain to better distinguish mental health domains, considering the target population.

Results

Of 4507 identified citations, 312 studies met the inclusion criteria, with 251 research articles and 61 trial protocols in registers.

Study characteristics, population, interventions, and methodology

Study populations differed across the studies, including heterogeneous populations with ID of unstated etiology ($n = 143$, 46%), GNDs ($n = 135$, 43%), and neurometabolic disorders ($n = 34$, 11%). Specific genetic or metabolic diagnoses that were included are presented in Table 1.

Sample sizes of identified trials ranged from 1 to 452 (median = 40) participants. Interventions included drug ($n = 123$, 39%), diet or supplement ($n = 14$, 4%), and non-drug interventions such as behavioral interventions ($n = 175$, 56%). Randomization was used in 224 (72%) of the studies. Studies were not blinded ($n = 155$, 50%), single-blinded ($n = 10$, 3%), double-blinded ($n = 85$, 27%), or blinding was unclear ($n = 62$, 20%).

In 7 (2%) of the clinical trials, it was explicitly mentioned that affected individuals or representatives were involved in the choice of outcome measures.

Reported outcomes

There were 438 different outcomes reported (Supplemental Material A), which we clustered into 91 different outcomes based on the most commonly used terminology (Table 2). Cognitive function was measured most frequently ($n = 333$, 28% of the measurements). Twenty-eight reported outcomes (31%) consisted of a combination of several outcomes, such as cognitive function and motor function.

Outcome measurement instruments

Of the 457 different outcome measurement instruments that were identified, 213 (47%) were classified as instruments for PROs, 54 (12%) as ClinROs, 48 (11%) as ObsROs, and 157 (34%)

Table 1. Number of clinical trials in ID of unstated etiology, genetic and neurometabolic disorders included in this review.

Diagnosis	*N*
ID of unstated etiology[a]	143
Down syndrome	33
Fragile X syndrome	23
Prader-Willi syndrome	23
Tuberous Sclerosis Complex	17
Mucopolysaccharidosis[b,c]	12
Neurofibromatosis type 1	9
Rett syndrome	8
Phenylketonuria[c]	7
Angelman syndrome	6
22q11.2 deletion syndrome	5
Niemann-Pick disease type C	4
Fragile X premutation-associated conditions	3
Smith-Magenis syndrome	3
1p36 deletion syndrome	2
Coffin-Siris syndrome	2
Cornelia de Lange syndrome	2
Kabuki syndrome	2
Metachromatic leukodystrophy	2
Phelan-McDermid syndrome	2
Succinic semialdehyde dehydrogenase deficiency	2
Williams syndrome	2
XYY syndrome[c]	2
Aicardi-Goutières Syndrome	1
Alpha-mannosidosis	1
Classic galactosemia	1
Congenital lipomatous overgrowth, vascular malformations, and epidermal nevi (syndrome)	1
Cyclin-dependent kinase like-5 deficiency disorder	1
Klippel-Trenaunay syndrome	1

(Continued)

Table 1. (Continued)

Diagnosis	*N*
Leigh syndrome	1
Mitochondrial disorders	1
Pantothemate kinase-associated neurodegeneration	1
Phosphomannomutase-2 congenital disorder of glycosylation	1
Propionic acidemia[c]	1
Smith-Lemli-Opitz syndrome	1

[a]One study included a participant with mucopolysaccharidosis type IV (Morquio A syndrome) which was however not considered associated with the diagnosed syndrome.
[b]Mucopolysaccharidosis type I ($n=3$), type II ($n=1$), type IIIA ($n=4$), type IIIB ($n=3$).
[c]Disorders that are not always associated with ID (i.e. due to advanced screening and therapies), but (some) participants included in these studies were affected with intellectual impairments.

as PerfOs. There were 288 (63%) outcome measurement instruments that were used in only one clinical trial. Another 16 (4%) outcome measurement instruments were self-designed for the particular trial, classified as instrument for PROs ($n=12$) and ObsROs ($n=4$). The outcome measurement instruments used to measure PROs, ClinROs, and ObsROs are reported in Supplemental Material B. Instruments for PerfOs measured cognitive function and physical function (Supplemental Material C). There were 18 condition-specific outcome measurement instruments used in 30 (10%) clinical trials in total, including instruments for Down syndrome, Prader-Willi syndrome, phenylketonuria, mitochondrial disease, Rett syndrome, Fragile X syndrome, Niemann-Pick disease type C, and phosphomannomutase deficiency congenital disorder of glycosylation (Supplemental Material B). Two condition-specific outcome measurement instruments were designed ad-hoc for the specific trial.

The outcome measurement instruments classified as PROs, ClinROs, and ObsROs were used as self-report ($n=183$, 22%), self-assisted report ($n=7$, 1%), proxy-report ($n=327$, 39%), observer-report ($n=46$, 6%), clinician-report ($n=132$, 16%) or unclear ($n=136$, 16%). Within proxy-report, parent-report was mentioned for 218 outcome measurements and teachers reported for 31 measurements (Supplemental Material B).

The instruments, classified according to the Valderas and Alonso model (when applicable), revealed representation of all health domains: symptoms ($n=26$, 5%), physical function ($n=34$, 7%), mental function ($n=141$, 29%), social function ($n=80$, 17%), general health perceptions ($n=16$, 3%), (HR-)QoL ($n=23$, 5%), and cognitive function including both performance-based tests and rating scales ($n=161$, 33%).

Discussion

This scoping review is the first overview of the myriad of outcomes and outcome measurement instruments used in clinical trials in GNDs and ID of unknown cause. It provides insight into the large number of (often differently reported) outcomes and measurement instruments. Cognitive function was most frequently measured. The majority of instruments was used in only one clinical trial. This review demonstrates the need for harmonization, consensus on terminology, classification, and development of a core outcome set. It serves as a starting point for discussion about a more universal approach to the selection of relevant outcomes and instruments, creating a bridge between GNDs and ID fields to enable evidence-based general ID care and measuring effectiveness of innovative therapies.

Reported outcomes

From a total of 312 studies, there were 438 different reported outcomes clustered into 91 different outcomes. We encountered differences in terminology for similar constructs, such as 'aberrant behaviors', 'challenging behaviors', 'behavioral problems', and 'severe behavioral manifestations'. This may conflict with generalizability and clarity among clinical trials, demonstrating the need for semantic harmonization. Similarly, overlap in PROs across the International Consortium for Health Outcomes Measurement standard sets was recently examined, identifying 307 different PROs referring to 22 unique PRO concepts.[37] Furthermore, (HR-)QoL was reported in 74 clinical trials, using 23 different instruments. Although HR-QoL is an important outcome, this broad, abstract, and multidimensional concept can cover different concepts, obscuring the

Table 2. Reported outcomes and number of outcome measurement instruments used, clustered according to frequency of use (for unclustered reported outcomes, see Supplemental Material A).

Reported outcomes	Frequency	Number of different outcome measurement instruments used	PRO	ClinRO	ObsRO	PerfO
Cognitive function	333	141	33	16	15	269
(HR-)QoL	74	23	72	2	0	0
Aberrant behavior [e.g. challenging/ maladaptive/dysfunctional/destructive behavior/(severe) behavioral problems/manifestations]	64	18	59	2	3	0
(Clinical) global impression (including severity/improvement)	64	16	15	49	0	0
Communication	59	35	11	0	13	35
Behavior (general/not specified)	45	23	36	3	6	0
Adaptive behavior	33	5	29	2	2	0
Depression and mood disorders	33	15	27	3	3	0
Autism	31	14	15	11	5	0
Anxiety	27	15	26	1	0	0
Mental health (e.g. global, well-being, feelings, psychological wellness/ distress, symptoms of mental disorder)	27	16	18	7	2	0
Social behavior	27	13	22	1	4	0
Sleep	26	12	19	0	7	0
Other/unclear	22	17	14	5	3	0
Motor function	20	19	3	7	2	8
Participation	18	11	13	0	5	0
Personalized goals	14	6	13	1	0	0
Emotion regulation	13	11	12	0	1	0
Activity	11	7	9	0	2	0
Syndrome-specific symptoms	10	8	5	5	0	0
Academic skills	9	9	2	0	1	6
Attention	9	9	5	1	3	0
Repetitive behavior	9	2	8	1	0	0
Aggression	9	1	1	0	5	0
Anger	8	4	8	0	0	0

(Continued)

Table 2. (Continued)

Reported outcomes	Frequency	Number of different outcome measurement instruments used	PRO	ClinRO	ObsRO	PerfO
Hyperphagia	8	1	7	1	0	0
Pain	7	6	7	0	0	0
Irritability	7	1	7	0	0	0
Self-injurious behavior	6	5	5	1	0	0
Eating behavior	5	4	5	0	0	0
Epilepsy	5	2	5	0	0	0
Psychiatric symptoms	5	3	1	4	0	0
Self-efficacy	5	5	3	1	1	0
Self-esteem	5	1	5	0	0	0
Ataxia	4	3	0	4	0	0
Neurological function	4	1	0	4	0	0
Social support	4	3	4	0	0	0
Stress	4	4	4	0	0	0
Substance use	4	3	4	0	0	0
Concerns	3	3	1	2	0	0
Psychosocial function	3	2	3	0	0	0
Post-traumatic stress disorder	3	2	3	0	0	0
Alertness	2	1	2	0	0	0
Coping behavior	2	2	2	0	0	0
Mentalizing abilities	2	2	0	0	0	2
Obsession and compulsivity	2	1	0	2	0	0
Resilience	2	2	2	0	0	0
Suicide	2	2	2	0	0	0
(Self-)compassion	1	1	1	0	0	0
Acceptance	1	1	1	0	0	0
Apathy	1	1	1	0	0	0
Confusion	1	1	1	0	0	0
Dysarthria	1	1	0	1	0	0
Dyskinesia	1	1	0	1	0	0
Dystonia	1	1	0	0	1	0

(Continued)

Table 2. (Continued)

Reported outcomes	Frequency	Number of different outcome measurement instruments used	PRO	ClinRO	ObsRO	PerfO
Empowerment	1	1	1	0	0	0
Hyperactivity	1	1	1	0	0	0
Life events	1	1	1	0	0	0
Psychosis	1	1	1	0	0	0
Satisfaction	1	1	1	0	0	0
Self-determination	1	1	1	0	0	0
Combined terms*						
Attention; hyperactivity; impulsivity	11	6	10	1	0	0
Anxiety; depression and mood disorders	7	2	7	0	0	0
Cognitive function; adaptive behavior	5	4	2	1	0	2
Cognitive function; motor function	5	4	1	0	2	2
Behavior; emotion regulation	4	2	4	0	0	0
Self-efficacy; social support	4	2	4	0	0	0
Cognitive function; motor function; communication	3	3	0	1	1	1
Behavior; cognitive function	2	1	0	2	0	0
Cognitive function; communication	2	2	0	0	0	2
Communication; activity; social behavior	2	1	2	0	0	0
Depression and mood disorders; behavior	2	1	2	0	0	0
Emotion regulation; social behavior	2	1	2	0	0	0
Emotion regulation; social behavior; eating behavior	2	1	2	0	0	0
Aggression; social behavior	1	1	1	0	0	0
Anxiety; irritability	1	1	1	0	0	0
Cognitive function; behavior	1	1	0	0	0	1
Cognitive function; Communication; social behavior	1	1	1	0	0	0
Cognitive function; Emotion regulation	1	1	0	0	0	1
Cognitive function; global impression	1	1	0	1	0	0

(Continued)

Table 2. (Continued)

Reported outcomes	Frequency	Number of different outcome measurement instruments used	PRO	ClinRO	ObsRO	PerfO
Cognitive function; motor function; emotion regulation	1	1	0	0	0	1
Communication; social behavior	1	1	1	0	0	0
Eating behavior; mental health	1	1	1	0	0	0
Emotion regulation; Social behavior; activity	1	1	1	0	0	0
Irritability; hyperactivity	1	1	1	0	0	0
Mental health; autism	1	1	0	1	0	0
Pain; mental health; social behavior	1	1	0	0	1	0
Pain; stress; social behavior	1	1	0	1	0	0
Satisfaction; mental health	1	1	1	0	0	0

*Combined terms include outcomes that are measured with one instrument, consisting of a combination of several outcomes.
ClinRO, clinician-reported outcome; ObsRO, observer-reported outcome; PerfO, performance outcome; PRO, patient-reported outcome.

construct to be measured. According to the FDA, an HR-QoL measure should at a minimum capture physical, psychological (including emotional and cognitive), and social functioning.[18]

Outcome measurement instruments

We identified 457 different outcome measurement instruments to measure PROs ($n=213$), ClinROs ($n=54$), ObsROs ($n=48$), and PerfOs ($n=157$), with 288 instruments (63%) only used in one clinical trial in the past decade. The large number of different instruments used in clinical trials is not surprising, considering the heterogeneity in levels of intellectual functioning, patients and researcher preferences, availability of instruments that are appropriate to specific conditions, and regional preferences. Furthermore, for novel drugs with yet unknown efficacy, multiple domains might be studied requiring different instruments, to investigate effectiveness and identify potential subgroups who benefit most from the intervention. This is also reflected by the large amount of ad-hoc designed symptom- specific and condition-specific instruments, hampering extrapolation and interpretation of the results. Yet, it is laborious to examine validity, reliability, and responsiveness of so many instruments. It underlines the need for consensus on outcomes and instruments, such as the Outcome Measures Working Groups and expert groups convened by the NIH,[38,39] the ERICA PROMs Repository (Endo-ERN), and the establishment and validation of the National Institutes of Health Toolbox Cognitive Battery (NIH-TCB) for individuals with ID.[28,40,41]

Type of reporter

Although instruments are generally developed as one specific type of COA, similar instruments were completed by different types of reporters (e.g. a ClinRO instrument used as ObsRO by parents). Furthermore, proxy-reports were substantially more used (39%) than self-(assisted) reports (23%). Although the use of proxy-reports is not surprising in populations with ID, the validity of proxy reflections of unobservable internal states (e.g. anxiety or depression) is limited, as the personal perspective can only truly be understood by the individual's self-report.[42] Proxy-raters often assess (HR-)QoL worse compared to individuals themselves, indicating bias.[43–49] PROs may thus be difficult to measure by proxy-reports,[50] although still providing valuable information.[51] It has been suggested that adolescents with ID can

reliably report on their mental health status, with instruments appropriate to their age, cognitive, and visuospatial functioning.[52,53] However, a recent study illustrated that there is currently no self-report instrument with acceptable psychometric properties available for assessing (HR-) QoL and subjective well-being of adolescents with ID.[42] Novel methods are upcoming such as experience sampling methods with the use of apps and ID-friendly instruments.[54,55]

The emergence of condition-specific and personalized outcome measurement instruments

In an attempt to measure the impact of a disease and target-specific phenotypes without the need for multiple tools,[56,57] condition-specific ($n=18$) and personalized ($n=6$) outcome measurement instruments were used. Also, tools particularly designed for a specific trial ($n=16$) were used. Condition-specific PROMs have also been developed due to unavailability of proxy-versions for adults with ID and criticism on appropriateness of the existing instrument's content and measurement properties for the target population.[56,58] Such instruments might contain more relevant items to complete, increasing acceptability among affected individuals. However, results might be difficult to generalize or interpret. Furthermore, it is not feasible and desirable to use condition-specific PROMs for more than 7000 rare disorders.[59] It may also not be necessary, as research has shown that PRO domains that patients consider important are very similar among patient populations.[37]

Generic instruments have the advantage of allowing comparison of outcomes between different disease (sub)groups. Generally, all individuals want to feel and function well, such as living without symptoms and being able to carry out daily activities. Feelings and functions can be affected by different health conditions, and these can result to similar problems with considerable overlap in relevant PROs across conditions, which could be measured with one set of generic outcome measures across conditions.[36,37] Methodological innovations, such as item response theory (IRT), have been used to develop PROMs with good measurement properties that are applicable across different health conditions, such as Patient-Reported Outcomes Measurement Information System (PROMIS®).[60–62] IRT-based item banks are large sets of calibrated questions measuring the same construct, enabling efficient measurement through short forms or computerized adaptive testing (CAT).[41,63] This provides a valuable solution, since redundant items for specific individuals will be minimized, increasing relevance and efficiency.[64]

To ensure relevance, personalized outcome measurement instruments have gained emerging interest, especially for rare and heterogeneous patient populations since health manifestations are often specific, variable, and complex.[65] Instruments such as Goal Attainment Scaling enable focusing on personal goals and abilities.[66] Additionally, by including outcomes that are specifically relevant to the affected individual, treatment adherence might be enhanced as well.[67] Also regulatory agencies have increasing interest in the relevance of what is being measured,[68] as treatment effects might be statistically significant, but not clinically or socially relevant, or vice versa.[69]

Recommendations for selecting outcomes and instruments

In order to measure what matters to patients, several factors should be taken into account when selecting outcome measurement instruments in clinical trials[70] (Table 3). First, it should be ensured that the construct being measured is relevant to the patient. Including relevant outcomes also contributes to recruitment and treatment compliance.[28,67] Affected individuals and representatives of the target population should be formally involved in the choice of measured outcomes, while now involvement was mentioned in only 2% of the clinical trials.

When selecting instruments, their acceptability, feasibility, and measurement properties should be taken into account, for example, by validation studies and using Consensus-based Standards for the Selection of Health Measurement Instruments criteria[71] (Table 3). For already over-burdened caregivers, outcome measurement instruments can be time-consuming to complete, and are often experienced as confronting due to inappropriateness of questions, leading to poor acceptability.[56] As such, IRT-based item banks could be used to enable efficient measurement through short forms or CAT.[41,63]

Table 3. Recommendations, as provided by the authors, with regard to selecting outcomes and outcome measurement instruments in clinical trials for individuals with GNDs and/or ID.

Considerations when selecting outcomes and instruments	Recommendations
What construct will be measured?	Make sure the construct is relevant to the affected individual(s) Formally involve affected individuals and/or representatives in the selection of measured outcomes
What instrument(s) could be used?	Take into account measurement properties, such as validity, reliability, and responsiveness to change Consider PROMIS®, core outcome sets, NIH-TCB, ERICA PROMs Repository
	Consider using different types of outcome measurement instruments, such as personalized measures, PROMs, and biological or mechanistic measures, which may also be relevant for translational research (e.g. measurable in animal studies) to enable comparison of candidate drugs across models and biomarkers
Is the instrument appropriate for this target population?	Take into account acceptability and feasibility to increase recruitment and compliance Minimize study visits and burden and maximize measurements in a natural setting (e.g. remote measurements and experience sampling methods)
Who will be the reporter?	Attempt to (also) acquire information directly from the affected individual, adapted to the level of functioning (e.g. smileys and other symbols)

ERICA, European Rare Disease Research Coordination and Support Action consortium; GND, genetic neurodevelopmental disorders; ID, intellectual disability; NIH-TCB, National Institutes of Health Toolbox Cognitive Battery; PROM, patient-reported outcome measure; PROMIS, Patient-Reported Outcomes Measurement Information System.

Furthermore, it is recommended to attempt to include (user-friendly) PROMs to acquire information from the patient perspective, as also encouraged by regulatory authorities such as the FDA and European Medicines Agency (EMA).[72,73]

Future outcome measure landscape

Because of the overgrowth of available outcome measurement instruments, clinical researchers need guidance in choosing appropriate outcome measures in clinical trials. Regulatory agencies, such as the EMA and FDA, encourage maintaining consistency in assessment methods and are placing focus on capturing the patient experience, but poorly defined PRO objectives have hindered the utility of PROs in regulatory decisions.[68,74] A core outcome set or generic measure with disorder-specific or comorbidity-specific extensions may provide a solution to ensure generalizability and interpretation, and effectively target-specific phenotypes in individuals with GNDs and ID. To move from this 'mess of measures' to a more unified approach for future interventional research for GNDs and ID, the field could take the following steps:

- Reach (international) consensus on outcomes (e.g. Delphi procedure) and establish a core outcome set for individuals with GNDs and ID: terminology and constructs should be relevant, clear, harmonized and operationalized, in collaboration with affected individuals, caregivers, and (methodological and clinical) experts.
- Reach (international) consensus on the most suitable instruments to be selected per outcome, taking into account relevance, applicability, patient preferences, validity, reliability, responsiveness to change, (strategies for controlling) learning effects, and

language and culture barriers. Some instruments may need to be adapted to individuals with ID.
- Implement the core outcome set or (ID-friendly) generic measure(s) with appropriate versions for different levels of ID. This could be extended with disorder- or comorbidity-specific measures (e.g. symptom checklists) to cover relevant condition-specific aspects.

Strengths and limitations

This scoping review is the first overview of outcomes and outcome measurement instruments used in clinical trials in GNDs and ID, examining the broad array of outcomes related to health manifestations common in these patient populations, using state of the art classifications. However, when conducting this review, we faced some challenges. We initially aimed to cluster the outcomes and outcome measurement instruments according to the Valderas and Alonso model.[12] Domain assessment has rather been an indication, as instruments should ideally be assessed per subscale (unidimensional), which was not feasible due to the enormous amount of different outcome measurement instruments. Furthermore, the terminology used for the outcomes and outcome measurement instruments was often unclear, lacking, or inconsistently reported. We clustered reported outcomes based on frequency of used terminology, and thus do not refer to a standardized terminology. Finally, we cannot recommend specific outcome measurement instruments, because psychometric properties were not investigated in this review.

Conclusion

This review provides insight into the large number of outcomes and outcome measurement instruments reported in clinical trials for GNDs and ID. The abundancy of available tools is problematic from an efficiency and generalizability perspective, highlighting the need for a more universal approach to the selection of outcomes and instruments. Moving forward, further collaborative efforts are recommended to achieve consensus on outcome selection. The output of this review may serve as a starting point for discussion about relevant outcomes and instruments in GNDs and ID, and to develop a core outcome set for these populations. Preferably, it will be applicable for care as well as research purposes with possible implications for market authorization and reimbursement of (orphan) drugs to improve patient-centered care by measuring what matters to affected individuals.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Author contributions

Annelieke R. Müller: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Visualization; Writing – original draft.

Nadia Y. van Silfhout: Conceptualization; Data curation; Formal analysis; Methodology; Writing – review & editing.

Bibiche den Hollander: Conceptualization; Data curation; Writing – review & editing.

Dick H. C. Kampman: Formal analysis; Investigation; Software; Writing – review & editing.

Lianne Bakkum: Conceptualization; Data curation; Methodology; Writing – review & editing.

Marion M. M. G. Brands: Conceptualization; Data curation; Methodology; Writing – review & editing.

Lotte Haverman: Conceptualization; Methodology.

Caroline B. Terwee: Methodology; Writing – review & editing.

Carlo Schuengel: Conceptualization; Methodology; Writing – review & editing.

Joost Daams: Conceptualization; Data curation; Writing – review & editing.

David Hessl: Methodology; Writing – review & editing.

Frits A. Wijburg: Conceptualization; Methodology; Supervision; Writing – review & editing.

Erik Boot: Conceptualization; Data curation; Methodology; Writing – review & editing.

Agnies M. van Eeghen: Conceptualization; Funding acquisition; Methodology; Supervision; Writing – review & editing.

Acknowledgement
None.

Funding
The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported financially by Stichting Wetenschappelijk Onderzoek 's Heeren Loo. The funder had no role in the design and conduct of the study, preparation of the review, or approval of the manuscript.

Competing interests
The authors declare that there is no conflict of interest.

Availability of data and materials
The search strategy is available on the Open Science Framework (https://osf.io/2zmxv/). The data extraction sheet is available on request.

ORCID iDs
Annelieke R. Müller https://orcid.org/0000-0002-2672-9989

Agnies M. van Eeghen https://orcid.org/0000-0001-8149-8645

Supplemental material
Supplemental material for this article is available online.

References

1. American Psychiatric Association. *American Psychiatric Association: diagnostic and Statistical Manual of Mental Disorders Fifth Edition*. Arlington, 2013.
2. Leonard H and Wen X. The epidemiology of mental retardation: challenges and opportunities in the new millennium. *Ment Retard Dev Disabil Res Rev* 2002; 8: 117–134.
3. Pinchefsky E and Shevell M. Intellectual disabilities and global developmental delay. In: Goldstein S and DeVries M (eds) *Handbook of DSM-5 Disorders in Children and Adolescents*, Springer Nature: Springer International Publishing, 2017, pp. 19–55.
4. Kvarnung M and Nordgren A. Intellectual disability & rare disorders: a diagnostic challenge. *Advances in Experimental Medicine and Biology*, 2017; 1031: 39–54.
5. Wang J, Wang Y, Wang L, *et al*. The diagnostic yield of intellectual disability: combined whole genome low-coverage sequencing and medical exome sequencing. *BMC Med Genomics* 2020; 13: 70.
6. Pekeles H, Accogli A, Boudrahem-Addour N, *et al*. Diagnostic yield of intellectual disability gene panels. *Pediatr Neurol* 2019; 92: 32–36.
7. Castrén E, Elgersma Y, Maffei L, *et al*. Treatment of neurodevelopmental disorders in adulthood. *J Neurosci* 2012; 32:14074–14079.
8. Levy G and Barak B. Postnatal therapeutic approaches in genetic neurodevelopmental disorders. *Neural Regen Res*. 2021; 16: 414–422.
9. Ferrari M. Borderline intellectual functioning and the intellectual disability construct. *Intellect Dev Disabil* 2009; 47: 386–389.
10. Swillen A and McDonald-Mcginn D. Developmental trajectories in 22q11.2 deletion syndrome. *Am J Med Genet Part C Semin Med Genet* 2015; 169: 172–181.
11. Green Snyder LA, D'Angelo D, Chen Q, *et al*. Autism spectrum disorder, developmental and psychiatric features in 16p11.2 duplication. *J Autism Dev Disord* 2016; 46: 2734–2748.
12. Valderas JM and Alonso J. Patient reported outcome measures: a model-based classification system for research and clinical practice. *Qual Life Res* 2008; 17:1125–1135.
13. Wilson IB. Linking clinical variables with health-related quality of life. *JAMA* 1995; 273; 59–65.
14. Cannizzo S, Lorenzoni V, Palla I, *et al*. Rare diseases under different levels of economic analysis: current activities, challenges and perspectives. *RMD Open* 2018; 4: e000794.
15. FDA Clinical Outcome Assessments, https://www.fda.gov/about-fda/cdrh-patient-science-and-engagement-program/clinical-outcome-assessments-coas-medical-device-decision-making (accessed 11 July 2023).
16. Gnanasakthy A, Qin S and Norcross L. FDA Guidance on selecting, developing, or modifying

fit-for-purpose clinical outcome assessments: old wine in a new bottle? *Patient* 2023; 16: 3–5.

17. Walton MK, Powers JH, Hobart J, *et al.* Clinical outcome assessments: conceptual foundation-report of the ISPOR clinical outcomes assessment-emerging good practices for outcomes research task force. *Value Heal* 2015; 18: 741–752.
18. U.S. Department of Health and Human Services FDA Center for Drug Evaluation and Research; U.S. Department of Health and Human Services FDA Center for Biologics Evaluation and Research; U.S. Department of Health and Human Services FDA Center for Devices and Radiological Health. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. *Health Qual Life Outcomes*. 2006; 4: 79.
19. Churruca K, Pomare C, Ellis LA, *et al.* Patient-reported outcome measures (PROMs): a review of generic and condition-specific measures and a discussion of trends and issues. *Health Expect* 2021; 24: 1015–1024.
20. Overwater IE, Rietman AB, van Eeghen AM, *et al.* Everolimus for the treatment of refractory seizures associated with tuberous sclerosis complex (TSC): current perspectives. *Ther Clin Risk Manag* 2019; 15: 951–955.
21. Van Eeghen AM, Bruining H, Wolf NI, *et al.* Personalized medicine for rare neurogenetic disorders: can we make it happen? *Cold Spring Harb Mol Case Stud* 2022; 8: a006200.
22. Jacquemont S, Berry-Kravis E, Hagerman R, *et al.* The challenges of clinical trials in fragile X syndrome. *Psychopharmacology* 2014; 231: 1237–1250.
23. Jansen-Van Der Weide MC, Gaasterland CMW, Roes KCB, *et al.* Rare disease registries: potential applications towards impact on development of new drug treatments. *Orphanet J Rare Dis* 2018; 13: 154.
24. Heussler H, Cohen J, Silove N, *et al.* A phase 1/2, open-label assessment of the safety, tolerability, and efficacy of transdermal cannabidiol (ZYN002) for the treatment of pediatric fragile X syndrome. *J Neurodev Disord* 2019; 11: 16.
25. Elgersma Y and Sonzogni M. UBE3A reinstatement as a disease-modifying therapy for Angelman syndrome. *Dev Med Child Neurol* 2021; 63: 802–807.
26. Northrup H, Aronow ME, Bebin EM, *et al.* Updated international tuberous sclerosis complex diagnostic criteria and surveillance and management recommendations. *Pediatr Neurol* 2021; 123: 50–66.
27. Müller AR, Brands MM, van de Ven PM, *et al.* The power of 1: systematic review of N-of-1 studies in rare genetic neurodevelopmental disorders. *Neurology* 2021; 96: 529–540.
28. Hessl D, Sansone SM, Berry-Kravis E, *et al.* The NIH Toolbox Cognitive Battery for intellectual disabilities: three preliminary studies and future directions. *J Neurodev Disord* 2016; 8: 35.
29. Berry-Kravis EM, Hessl D, Rathmell B, *et al.* Effects of STX209 (Arbaclofen) on neurobehavioral function in children and adults with fragile X syndrome: a randomized, controlled, phase 2 trial. *Sci Transl Med* 2012; 4:152ra127.
30. Berry-Kravis E, Hagerman R, Visootsak J, *et al.* Arbaclofen in fragile X syndrome: results of phase 3 trials. *J Neurodev Disord* 2017; 9: 3.
31. Erickson CA, Davenport MH, Schaefer TL, *et al.* Fragile X targeted pharmacotherapy: lessons learned and future directions. *J Neurodev Disord* 2017; 9: 7.
32. Tricco AC, Lillie E, Zarin W, *et al.* PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med* 2018; 169: 467–473.
33. Moher D, Shamseer L, Clarke M, *et al.* Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev* 2015; 4(1): 1.
34. Poletti V and Biffi A. Gene-based approaches to inherited neurometabolic diseases. *Hum Gene Ther* 2019; 30: 1222–1235.
35. Ouzzani M, Hammady H, Fedorowicz Z, *et al.* Rayyan-a web and mobile app for systematic reviews. *Syst Rev* 2016; 5: 210.
36. Dodd S, Clarke M, Becker L, *et al.* A taxonomy has been developed for outcomes in medical research to help improve knowledge discovery. *J Clin Epidemiol* 2018; 96: 84–92.
37. Terwee CB, Zuidgeest M, Vonkeman HE, *et al.* Common patient-reported outcomes across ICHOM Standard Sets: the potential contribution of PROMIS®. *BMC Med Inform Decis Mak* 2021; 21: 259.
38. Budimirovic DB, Berry-Kravis E, Erickson CA, *et al.* Updated report on tools to measure outcomes of clinical trials in fragile X syndrome. *J Neurodevelop Disord* 2017; 9: 14.
39. Esbensen AJ, Hooper SR, Fidler D, *et al.* Outcome measures for clinical trials in down

syndrome. *Am J Intellect Dev Disabil* 2017; 122: 247–281.

40. Endo-ERN. The ERICA Patient Reported Outcome Measures (PROMs) Repository, https://endo-ern.eu/the-erica-patient-reported-outcome-measures-proms-repository/ (accessed 3 June 2023).
41. Shields RH, Kaat AJ, McKenzie FJ, *et al.* Validation of the NIH Toolbox Cognitive Battery in intellectual disability. *Neurology* 2020; 94: e1229–e1240.
42. Maguire S, Davison J, McLaughlin M, *et al.* Exploring the psychometric properties of self-report instruments used to measure health-related quality of life and subjective wellbeing of adolescents with intellectual disabilities: a Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) systematic review. *J Appl Res Intellect Disabil.* 2023; 36: 899–915. http://www.ncbi.nlm.nih.gov/pubmed/37101341
43. Burks HB, Des Bordes JKA, Chadha R, *et al.* Quality of life assessment in older adults with dementia: a systematic review. *Dement Geriatr Cogn Disord* 2021; 50: 103–110.
44. Griffiths AW, Smith SJ, Martin A, *et al.* Exploring self-report and proxy-report quality-of-life measures for people living with dementia in care homes. *Qual Life Res* 2020; 29: 463–472.
45. Alcantara J, Ohm J and Alcantara J. Comparison of pediatric self reports and parent proxy reports utilizing PROMIS: results from a chiropractic practice-based research network. *Complement Ther Clin Pract* 2017; 29: 48–52.
46. Janse AJ, Gemke RJBJ, Uiterwaal CSPM, *et al.* Quality of life: patients and doctors don't always agree: a meta-analysis. *J Clin Epidemiol* 2004; 57: 653–661.
47. Lunsky Y and Bramston P. A preliminary study of perceived stress in adults with intellectual disabilities according to self-report and informant ratings. *J Intellect Dev Disabil* 2006; 31: 20–27.
48. Morrow AM, Hayen A, Quine S, *et al.* A comparison of doctors', parents' and children's reports of health states and health-related quality of life in children with chronic conditions. *Child Care Health Dev* 2012; 38: 186–195.
49. Janssen CGC, Schuengel C and Stolk J. Perspectives on quality of life of people with intellectual disabilities: the interpretation of discrepancies between clients and caregivers. *Qual Life Res* 2005; 14: 57–69.
50. Ediebah DE, Reijneveld JC, Taphoorn MJB, *et al.* Impact of neurocognitive deficits on patient–proxy agreement regarding health-related quality of life in low-grade glioma patients. *Qual Life Res* 2017; 26: 869–880.
51. Upton P, Lawford J and Eiser C. Parent–child agreement across child health-related quality of life instruments: a review of the literature. *Qual Life Res* 2008; 17: 895–913.
52. Davison J, Maguire S, McLaughlin M, *et al.* Involving adolescents with intellectual disability in the adaptation of self-reported subjective well-being measures: participatory research and methodological considerations. *J Intellect Disabil Res* 2022; 66: 628–641.
53. Ingerski LM, Modi AC, Hood KK, *et al.* Health-related quality of life across pediatric chronic conditions. *J Pediatr* 2010; 156: 639–644.
54. Kooijmans R, Mercera G, Langdon PE, *et al.* The adaptation of self-report measures to the needs of people with intellectual disabilities: a systematic review. *Clin Psychol Sci Pract* 2022; 29: 250.
55. Bakkum L, Paalman C, Müller AR, *et al.* Accessibility and feasibility of experience sampling methods for mental health research with people with intellectual disability: scoping of research and stakeholder views. *J Appl Res Intellect Disabil* 2024; 37: e13190.
56. Kang S, Jones A, Shaffer RC, *et al.* Developing improved outcome measures in FXS: key stakeholder feedback. *Res Develop Disabil* 2023; 137: 104502.
57. Adang LA, Gavazzi F, Jawad AF, *et al.* Development of a neurologic severity scale for Aicardi Goutières Syndrome. *Mol Genet Metab* 2020; 130: 153–160.
58. Müller AR, Luijten MAJ, Haverman L, *et al.* Understanding the impact of tuberous sclerosis complex: development and validation of the TSC-PROM. *BMC Med* 2023; 21: 298.
59. Rode J. *Rare diseases: understanding this public health priority*. Paris: EURORDIS, 2005.
60. Cella D, Yount S, Rothrock N, *et al.* The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH roadmap cooperative group during its first two years. *Med Care* 2007; 45(Suppl. 1): S3–S11.
61. Cella D, Riley W, Stone A, *et al.* The patient-reported outcomes measurement information system (PROMIS) developed and tested its first

wave of adult self-reported health outcome item banks: 2005–2008. *J Clin Epidemiol* 2010; 63: 1179–1194.

62. Nguyen TH, Han HR, Kim MT, *et al*. An introduction to item response theory for patient-reported outcome measurement. *Patient* 2014; 7: 23–35.
63. Luijten MAJ, van Litsenburg RRL, Terwee CB, *et al*. Psychometric properties of the Patient-Reported Outcomes Measurement Information System (PROMIS®) pediatric item bank peer relationships in the Dutch general population. *Qual Life Res* 2021; 30:2061–2070.
64. Fries JF, Witter J, Rose M, *et al*. Item response theory, computerized adaptive testing, and promis: assessment of physical function. *J Rheumatol* 2014; 41:153–158.
65. Gaasterland CMW, Van Der Weide MCJ, Roes KCB, *et al*. Goal attainment scaling as an outcome measure in rare disease trials: a conceptual proposal for validation. *BMC Med Res Methodol* 2019; 19: 227.
66. Müller AR, Zinkstok JR, Rommelse NNJ, *et al*. Methylphenidate for attention-deficit/hyperactivity disorder in patients with Smith–Magenis syndrome: protocol for a series of N-of-1 trials. *Orphanet J Rare Dis* 2021; 16: 380.
67. Gaasterland CMW, Jansen-van der Weide MC, Vroom E, *et al*. The POWER-tool: recommendations for involving patient representatives in choosing relevant outcome measures during rare disease clinical trial design. *Health Policy (New York)*. 2018; 122: 1287–1294.
68. Crossnohere NL, Brundage M, Calvert MJ, *et al*. International guidance on the selection of patient-reported outcome measures in clinical trials: a review. *Qual Life Res* 2021; 30: 21–40.
69. Pogrow S. How Effect size (practical significance) misleads clinical practice: the case for switching to practical benefit to assess applied research findings. *Am Stat* 2019; 73(Suppl. 1): 223–234.
70. Morel T and Cano SJ. Measuring what matters to rare disease patients – Reflections on the work by the IRDiRC taskforce on patient-centered outcome measures. *Orphanet J Rare Dis* 2017;12: 171.
71. Terwee CB, Prinsen CAC, Chiarotto A, *et al*. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res* 2018; 27: 1159–1170.
72. Matts ST, Webber CM, Bocell FD, *et al*. Inclusion of patient-reported outcome instruments in US FDA medical device marketing authorizations. *J Patient-Rep Outcomes* 2022; 6: 38.
73. US Food and Drug Administration. *Guidance for Industry Patient-Reported Outcome Measures Use in Medical Product Development to Support Labeling Claims*. Clinical/Medical Federal Register, 2009.
74. Calvert M, King M, Mercieca-Bebber R, *et al*. SPIRIT-PRO Extension explanation and elaboration: guidelines for inclusion of patient-reported outcomes in protocols of clinical trials. *BMJ Open* 2021; 11: e045105.