Systematic Review of N-of-1 Studies in Rare Genetic Neurodevelopmental Disorders

The Power of 1

Annelieke R. Müller, MSc, Marion M.M.G. Brands, MD, PhD, Peter M. van de Ven, PhD, Kit C.B. Roes, PhD,
Martina C. Cornel, MD, PhD, Clara D.M. van Karnebeek, MD, PhD, Frits A. Wijburg, MD, PhD,
Joost G. Daams, MA, Erik Boot, MD, PhD, and Agnies M. van Eeghen, MD, PhD

Correspondence
Dr. van Eeghen
a.m.vaneeghen@
amsterdamumc.nl

Abstract

Objective
To improve the use of N-of-1 studies in rare genetic neurodevelopmental disorders, we systematically reviewed the literature and formulated recommendations for future studies.

Methods
The systematic review protocol was registered in the PROSPERO International Prospective Register of Systematic Reviews (CRD42020154720). EMBASE and MEDLINE were searched for relevant studies. Information was recorded on types of interventions, outcome measures, validity, strengths, and limitations using standard reporting guidelines and critical appraisal tools. Qualitative and descriptive analyses were performed.

Results
Twelve studies met the N-of-1 inclusion criteria, including both single trials and series. Interventions were mainly directed to neuropsychiatric manifestations. Main strengths were the use of personalized and clinically relevant outcomes in most studies. Generalizability was compromised due to limited use of validated and generalizable outcome measures.

Conclusion
N-of-1 studies are sporadically reported in rare genetic neurodevelopmental disorders. Properly executed N-of-1 studies may provide a powerful alternative to larger randomized controlled trials in rare disorders and a much needed bridge between practice and science. We provide recommendations for future N-of-1 studies in rare genetic neurodevelopmental disorders, ultimately optimizing evidence-based and personalized care.

's Heeren Loo (A.R.M.), Amersfoort, the Netherlands, and Amsterdam UMC (A.R.M.), Pediatric Metabolic Diseases, Emma Children's Hospital, University of Amsterdam, Amsterdam, the Netherlands; Pediatric Metabolic Diseases (M.M.G.B), Amsterdam UMC, Emma Children's Hospital, University of Amsterdam, Amsterdam, the Netherlands; Department of Epidemiology and Biostatistics (P.M.v.d.V.), Amsterdam UMC, Amsterdam, the Netherlands; Department of Health Evidence, Biostatistics (K.C.B.R.), Radboud University Medical Center, Radboud University, Nijmegen, the Netherlands; Department of Clinical Genetics (M.C.C.), Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam Public Health Research Institute, Amsterdam, the Netherlands; Pediatric Metabolic Diseases (C.D.M.v.K.), Amsterdam UMC, Emma Children's Hospital, University of Amsterdam, Amsterdam, the Netherlands, and Department of Pediatrics (C.D.M.v.K.), Radboud University Medical Center, Radboud Centre for Mitochondrial Medicine, Nijmegen, the Netherlands; Pediatric Metabolic Diseases (F.A.W.), Amsterdam UMC, Emma Children's Hospital, University of Amsterdam, Amsterdam, the Netherlands; Medical Library (J.G.D.), Amsterdam UMC, University of Amsterdam, Amsterdam, the Netherlands; 's Heeren Loo (E.B.), Amersfoort, the Netherlands, and Department of Psychiatry and Neuropsychology (E.B.), Maastricht University, Maastricht, the Netherlands, University Health Network (E.B.), The Dalglish Family 22q Clinic, Toronto, Ontario, Canada; and 's Heeren Loo (A.M.v.E.), Amersfoort, the Netherlands, Amsterdam UMC (A.M.v.E.), Emma Children's Hospital, University of Amsterdam, Amsterdam, the Netherlands, and Erasmus Medical Center (A.M.v.E.), ENCORE, Rotterdam, the Netherlands.

Glossary

ID = intellectual disability; **IEM** = inborn errors of metabolism; **RCT** = randomized controlled trial; **RoBiNT** = Risk of Bias in N-of-1 Trial.

Millions of people worldwide are affected by one of the nearly 8,000 rare disorders, defined as a condition affecting less than 1 in 2,000 individuals according to European definitions.[1] Around 80% of these rare disorders are genetic and associated with neurodevelopmental disorders and/or inborn errors of metabolism (IEMs).[2] Treatment targets are increasingly identified,[3,4] although the lack of evidence now leads to patients missing out on possibly effective interventions. As parallel group randomized controlled trials (RCTs) are often not feasible in these small and heterogeneous populations, a new methodological framework needs to be developed.

N-of-1 studies are randomized, controlled, multiple crossover trials in a single patient (figure 1 and table 1)[5,6] and closely follow indications of causality between agent and effect.[7,8] Where RCTs generally assess an average treatment effect, N-of-1 series identify individual particular characteristics that may modify response to the intervention, addressing the question of interindividual variability in treatment response.[9] Aggregated data of an N-of-1 series can even produce treatment effect estimates at a population level, which may be as robust as traditional RCTs.[10,11] Furthermore, the personalized approach has the potential of maximizing treatment adherence.[5,12–14]

Now guidelines on the design and reporting of N-of-1 trials are available,[6,8,15,16] and specific information is needed to improve N-of-1 studies in patients with rare neurodevelopmental disorders, as these patient populations are particularly complex, heterogeneous, vulnerable, and understudied. Our aim is to (1) provide a systematic review of the literature on N-of-1 trials in individuals with rare genetic neurodevelopmental disorders and (2) formulate recommendations to optimize future use and impact.

Methods

We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analysis Protocol (figure 2).[17] The methodological framework was published in advance in the PROSPERO International Prospective Register of Systematic Reviews (CRD42020154720). Relevant definitions of terms that were used in this review are provided in the box.

Eligibility

Peer-reviewed studies that used at least 3 controlled episodes of treatment or comparator (placebo, treatment as usual, no intervention, an alternative intervention, or other doses of the same intervention) were included in the review. Genetic neurodevelopmental disorders were defined as disorders with a genetic etiology affecting the nervous system in early development. IEMs, constituting a subgroup of rare genetic disorders, were defined as monogenic conditions in which the impairment of a biochemical pathway is essential to the pathophysiology of the disorder, typically resulting in either accumulation of toxic metabolites or shortage of energy and building blocks for cells. Those presented with intellectual disability (ID) were considered neurodevelopmental.[18,19] Exclusion criteria included idiopathic psychiatric disorders according to the DSM-5 criteria and genetic etiologies not confirmed with standard methods. Experts were consulted to determine whether the phenotypes of Rett syndrome (in the absence of molecular confirmation) and cerebellar hypoplasia tapetoretinal degeneration syndrome were consistent with the tight diagnosis.[20,21]

Search Strategy, Study Selection, Risk of Bias, and Quality Assessment

Two separate search strategies were conducted with assistance of a clinical research librarian (J.G.D.) in 2 search engines: MEDLINE (Ovid), 1946 to November 8, 2019, and EMBASE (Ovid), 1947 to November 8, 2019. First, the term N-of-1 and synonyms for all single-case experimental designs were searched. Second, because few studies explicitly used this N-of-1 terminology, all rare genetic neurodevelopmental disorders were separately searched in combination with terms for clinical trials. Specifically, a list containing all rare genetic and chromosome disorders and IEMs from the Genetic and Rare Diseases Information Center of the NIH was used. A time limit for the second search strategy in EMBASE of the last 10 years was applied due to the large amount of articles. Additional articles were identified by scoping search (n = 15), reference list checking and citation tracking (n = 59), and contacting authors of relevant articles (n = 6). All searches were conducted by the librarian and 1 reviewer (A.R.M.).

Rayyan (an application for systematic reviews) was used for screening.[22] All titles and abstracts were screened for relevance by 4 reviewers (A.M.v.E., M.M.M.G.B., E.B., and A.R.M.) with a subsample of 10% screened for interrater reliability. Interrater reliability analysis using the Cohen kappa statistic was performed to determine consistency between raters. Full texts were screened against inclusion and exclusion criteria, and data were independently extracted by at least 2 reviewers, of whom 1 (A.R.M.) covering all studies. Discrepancies were discussed until consensus was reached.

To provide guidance for appraisal of the quality of reporting of the full text publications and methodology, the Consolidated Standards of Reporting Trials (CONSORT) extension for

Figure 1 Schematic Presentation of Terminology Used for N-of-1 Trials
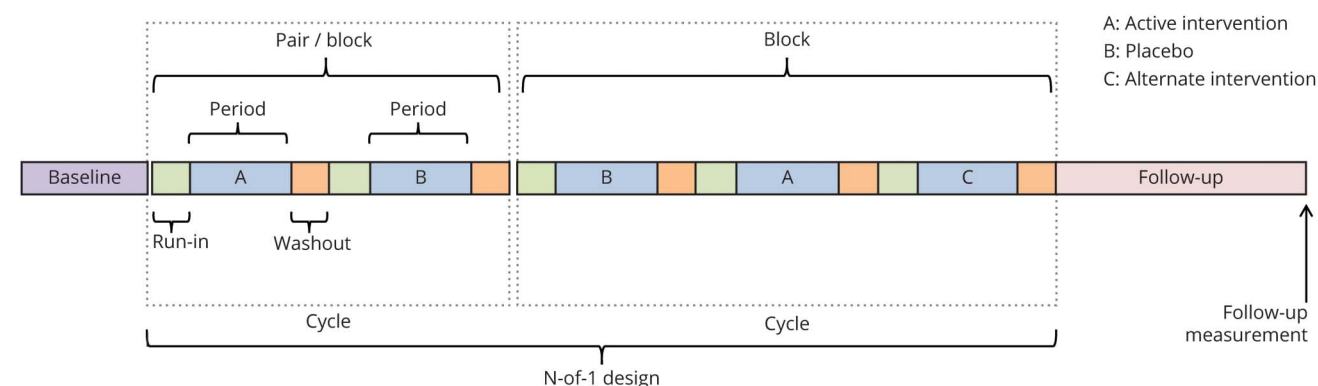
reporting N-of-1 Trials (CENT) 2015[6,23] and the Risk of Bias in N-of-1 Trials (RoBiNT) Scale[15] were scored. The CENT 2015 reporting standard consists of 25 items including recommendations about what to report and covers optimal methodology of medical and behavioral sciences. The RoBiNT Scale consists of 15 items including subscales on internal and external validity and evaluates how well a particular component of a study is conducted. The internal validity scale of the RoBiNT consists of 7, and the external validity and interpretation scale of 8 items, with a maximum score of 14 and 16 points, respectively.
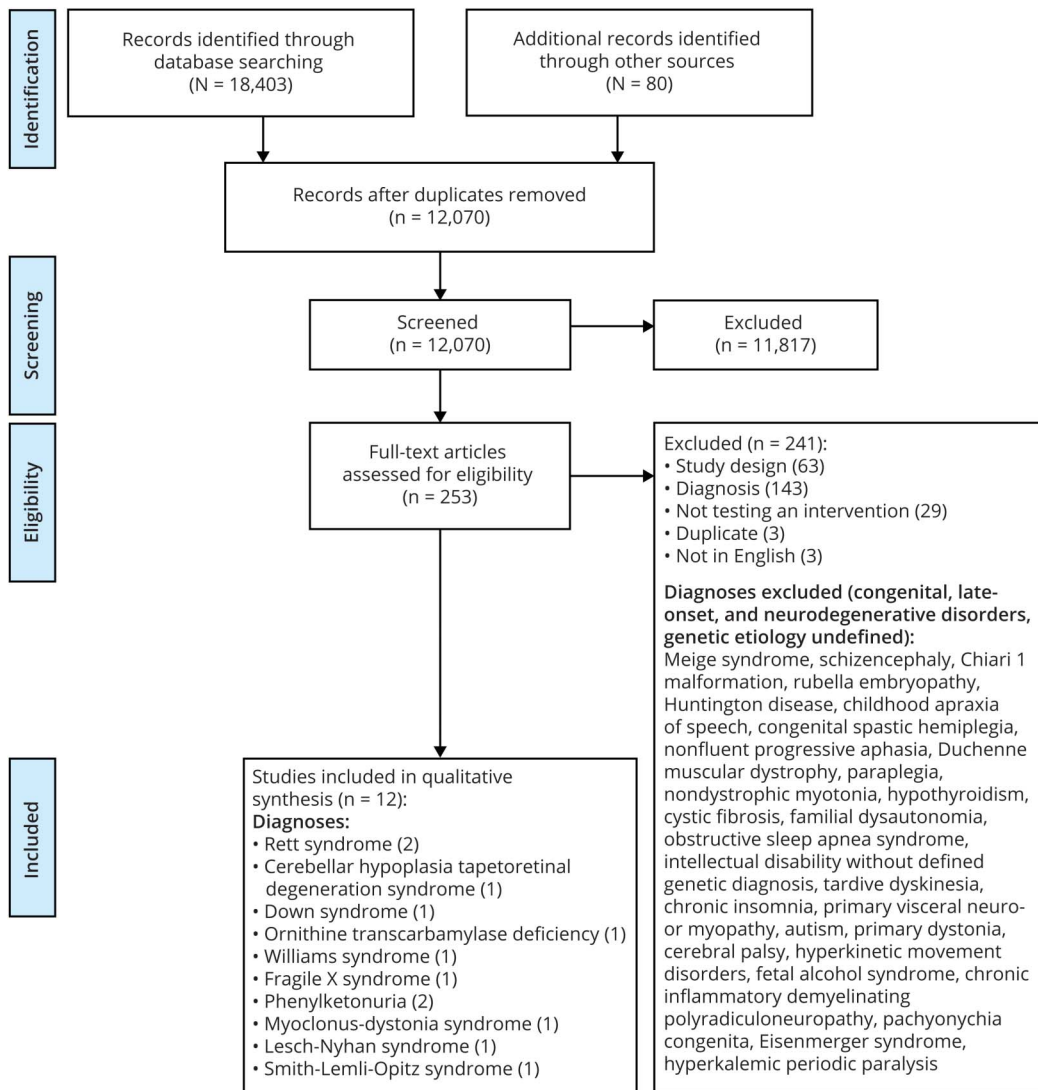
Data Extraction

Data were extracted on first author, year of publication, countries of study, number of participants, diagnosis, patient characteristics (age, presence/absence of ID, level of ID, Full Scale Intelligence Quotient, psychiatric diagnosis, comorbidities, and concurrent therapies), selection criteria, institutional ethical approval, trial design, run-in and washout periods, number of trial conditions, number and duration of periods, randomization, blinding, crossover trials, intervention(s), total intervention duration, comparator used, outcome assessment, major organ system studied, primary/secondary

Table 1 N-of-1 Methodological Terminology

Adherence	The extent to which a patient's behavior matches agreed recommendations from a health care provider taking into account the patient's perspectives.
Block	A repeated unit of a set number of periods.
Compliance	The extent to which the patient's behavior corresponds with the prescriber's recommendations.
Cycle	Each repeated unit of a set number of periods within a sequence (e.g., ABA).
Generalization	The degree to which results observed in a study may extend to other patients or settings, providing an indication of the external validity.
Generalization measure	Dependent variables in addition to the target behavior used to evaluate transfer effects of the intervention to a broader domain of functioning including other behaviors or settings.
Internal validity	The degree to which the study's outcomes could be attributed to the intervention being responsible for change in the dependent variable.
N-of-1 study	A prospectively planned randomized, controlled multiple crossover trial to determine the effectiveness of an intervention (A) in a single participant. Comparators (B) may include placebo, usual care, alternate treatment, or no intervention.
Pair	A repeated unit containing only 2 periods.
Period	The duration of an intervention, comparator, washout, or run-in.
Suggested inference	Interpretation of the extent of generalization of the study's outcomes to either the individual participants or patients in general with that specific disorder.
Responsiveness to change	The ability of an instrument to detect change over time in a construct being measured.
Run-in	Time preceding starting treatment at intended dose to avoid sudden introduction of a fixed therapeutic dose to determine participant compliance or to wash out effects of a previous drug.
Washout	Time without an intervention following a treatment period to ensure that effects of treatment have disappeared.

Figure 2 PRISMA Flowchart


```
Identification
┌─────────────────────────┐   ┌─────────────────────────┐
│ Records identified through│   │ Additional records identified│
│ database searching        │   │ through other sources     │
│ (N = 18,403)             │   │ (N = 80)                 │
└─────────────────────────┘   └─────────────────────────┘
```

Records after duplicates removed
(n = 12,070)

Screening

Screened
(n = 12,070)

Excluded
(n = 11,817)

Eligibility

Full-text articles
assessed for eligibility
(n = 253)

Excluded (n = 241):
• Study design (63)
• Diagnosis (143)
• Not testing an intervention (29)
• Duplicate (3)
• Not in English (3)

Diagnoses excluded (congenital, late-onset, and neurodegenerative disorders, genetic etiology undefined):
Meige syndrome, schizencephaly, Chiari 1 malformation, rubella embryopathy, Huntington disease, childhood apraxia of speech, congenital spastic hemiplegia, nonfluent progressive aphasia, Duchenne muscular dystrophy, paraplegia, nondystrophic myotonia, hypothyroidism, cystic fibrosis, familial dysautonomia, obstructive sleep apnea syndrome, intellectual disability without defined genetic diagnosis, tardive dyskinesia, chronic insomnia, primary visceral neuro- or myopathy, autism, primary dystonia, cerebral palsy, hyperkinetic movement disorders, fetal alcohol syndrome, chronic inflammatory demyelinating polyradiculoneuropathy, pachyonychia congenita, Eisenmerger syndrome, hyperkalemic periodic paralysis

Included

Studies included in qualitative
synthesis (n = 12):
Diagnoses:
• Rett syndrome (2)
• Cerebellar hypoplasia tapetoretinal
  degeneration syndrome (1)
• Down syndrome (1)
• Ornithine transcarbamylase deficiency (1)
• Williams syndrome (1)
• Fragile X syndrome (1)
• Phenylketonuria (2)
• Myoclonus-dystonia syndrome (1)
• Lesch-Nyhan syndrome (1)
• Smith-Lemli-Opitz syndrome (1)

PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analysis.

outcome measure(s) (presence and type), adverse events, power analysis, method of primary and/or secondary analysis (qualitative, graphical, tabular, (non)parametric statistics, and Bayesian statistics), main results, suggested inference, and challenges. The interventions were classified into disease-modifying or symptomatic, using a standard definition of disease-modifying: an intervention mediating the effect by targeting the primary underlying pathophysiology and changing the course of the disease with an enduring effect.[24] The suggested inference appraised generalization of the study's outcomes to either the individual participants or patients in general with that specific disorder. Generalization measures were defined as dependent variables in addition to the target behavior used to evaluate transfer effects of the intervention to a broader domain of functioning including other behaviors or settings.[15] A generalization measure could be an assessment of the same behavior in different settings or a measurement of an interventional effect on a completely different behavior. These should be identified a priori and measured throughout all phases. Strengths, limitations, and recommendations noted by the author(s) were collected, and reviewers were asked for additional comments.

Data Availability
The search strategy and data extraction sheet are available on request to the first author.

Results

Of 18.483 identified citations, 12 studies met the inclusion criteria, summarized in table 2. One article reported on 2 different N-of-1 studies with divergent methodological characteristics.[25]

Study Characteristics
Institutional ethics approval was explicitly mentioned in 8 studies.

Table 2 Characteristics of N-of-1 Studies in Rare Genetic Neurodevelopmental Disorders

Study	Diagnosis	No. of participants	Average age of participants (range); y	Intervention	Primary and secondary outcome measures[a]	Assessed by
Bawden et al.[55]	Williams syndrome	4	11 (9–13)	Methylphenidate	Child Behavior Checklist, Conners Parent/Teacher Questionnaire, Side Effects Questionnaire, and cognitive psychometric measures	Caregiver
Byiers et al.[56]	Rett syndrome	3	30 (15–47)	Functional communication training	*Communicative behavior*	Investigator
Camfield et al.[21]	Cerebellar hypoplasia tapetoretinal degeneration syndrome	6	7 (3–13)	Melatonin	Average number of hours asleep per 24 h and the number of awakenings and nights without arousals	Caregiver and parents
Crook et al.[57]	Down syndrome	5	59 (55–63)	Cognitive stimulation therapy	*Dementia Care Mapping*	Caregiver
Fisch et al.[26]	Fragile X syndrome	6	8 (3–15)	Folic acid	*Vineland Adaptive Behavior Scales, Autistic Descriptors Checklist*, questionnaire about noticed changes in behavior,and red blood cell folate levels	Caregiver and parents
Giffin et al.[27]	Phenylketonuria	3	15 (9–21)	Phenylalanine restriction	*Visual attention*, plasma phenylalanine, and tyrosine levels	Investigator
Hackett et al.[28]	Ornithine transcarbamylase deficiency	1	48	ʟ-arginine	*Quality of life/mood assessment questionnaire*, plasma glutamine, and arginine levels	Patient and investigator
Khasnavis et al.[30]	Lesch-Nyhan disease	9	10 (6–22)	Ecopipam	*Behavior Problems Inventory*, Clinical Global Impression scale, and adverse events	Caregiver and study staff
Luciano et al.[29]	Myoclonus-dystonia syndrome	2	29 (28–31)	Tetrabenazine	Global Dystonia rating scale, Fahn-Marsden rating scale, and Unified Myoclonus Rating Scale	Investigator
Marholin et al.[25]	Phenylketonuria	6	36 (19–53)	Low phenylalanine diet and behavior modification	Social and motor behavior and serum phenylalanine levels	Investigator
Simacek et al.[20]	Rett syndrome	3	3 (3–4)	Functional communication training	*Idiosyncratic responses and augmentative and alternative communication requests*	Investigator
Tierney et al.[31]	Smith-Lemli-Opitz syndrome	10	11 (5–20)	Cholesterol—easy eggs liquid egg yolks	*Aberrant Behavior Checklist (ABC)*	Caregiver

[a] Italics when indicated as a primary outcome measure by the authors.

Population

The 12 included studies had an average of 5 participants with an average age of 21 (range 3–63) years (table 2). The majority of the studies (n = 7) did not define the eligibility criteria.

Intervention

Various types of interventions were applied: psychological therapy (n = 4), dietary supplement (n = 4), drug (n = 3), and dietary therapy (n = 2; table 2). One study combined 2 subsequent interventions.[25] Only some dietary interventions might be categorized as disease modifying including phenylalanine restriction and folic acid and ʟ-arginine supplementation,[25–28] although distinction was difficult due to vague demarcations in targeting the possibly underlying mechanisms. Concurrent therapies were mentioned in 7 studies.

Methodological Characteristics

There was a wide variety of methodological approaches in the reviewed studies with great variation in number of periods and trial conditions and duration of the interventional period (table 3). Only 1 study included a washout period, and 1 study a run-in. Randomization was applied in 7 studies. None of those that did randomize explicitly specified the method of randomization. Seven studies were double blinded, 2 single blinded, and 4 were not blinded. The main comparator used

Table 3 Methodological Characteristics of the 12 Included N-of-1 Studies

Study (first author)	Design	Periods	Duration active intervention	Total trial duration	Randomization	Blinding	Comparator(s)	Type of analyses
Bawden	ABAB/ABBA/BAAB	5	1 wk	5 wks	Yes	Double blind	Placebo	Tabular
Byiers	ABAB	4	15–35 mins	2–3 ds	No	Not blinded	Alternate treatment	Graphical, nonoverlap of all pairs
Camfield	ABABABABAB	10	1 wk	10 wks	Yes	Double blind	Placebo	Tabular
Crook	AA[AAABBBCCC]	11	30 mins	11 d	Yes	Not blinded	Alternate treatment + no intervention	Graphical, tabular, and statistics ((non) parametric)
Fisch	ABA/BAB	3	4 mos	12 mos	Yes	Double blind	Placebo	Graphical and tabular
Giffin	ABA	3	8–14 wks	16–26 wks	No	Patient and observer	Alternate treatment	Statistics ((non) parametric) and graphical
Hackett	ABABAB	6	1 wk	7 wks	Yes	Double blind	Placebo	Statistics ((non) parametric) and tabular
Khasnavis	ABA/BAB + A (follow-up) + open-label extension	5	6 wks	17 mos	Yes	Double blind	Placebo + no intervention	Tabular and graphical
Luciano	ABAB	4	<1 d	3 d	No	Clinician only	No intervention	Tabular
Marholin	ABA	3	56 d (range 53–59)	Not reported	No	Double blind	No intervention	Graphical
	ABCADAD	7	2–11 d	45 d	No	Not blinded	No intervention	Graphical
Simacek	ABAB	8	5 mins	Not reported	No	Not blinded	Treatment as usual	Graphical
Tierney	ABACA/ACABA (with washout)	5	2 wks	10 wks	Yes	Double blind	Placebo + treatment as usual	Statistics ((non-) parametric) and graphical

was placebo followed by no intervention, with some studies applying a combination of several comparators. Graphical or tabular analyses were most often used to assess treatment effects. In 4 studies, (non)parametric statistical analyses were performed.

Outcome Measures and Evaluation Methods

In 9 studies, a primary outcome measure was present and predefined, although only 3 studies explicitly used the term primary outcome measure. Generally, outcome measures were targeted at behavioral and cognitive improvements (table 2). The evaluation methods used were diverse, varying from validated questionnaires to self-designed scoring lists. Only in myoclonus-dystonia syndrome, condition-specific rating scales were used.[29] Once, a quality of life assessment was used.[28] In 4 studies, biological plasma measurements were assessed to confirm an appropriate blood level of either the supplement or diet. None of the studies included generalization measures. Mostly, outcomes were assessed by caregivers and to a lesser degree by investigators.

Main Results and Adverse Events

Neither the supplement nor diet interventions revealed significant positive results, whereas results of drug interventions varied and nondrug interventional studies all reported positive effects, though not substantiated with statistical analysis. One study had to be prematurely discontinued due to unexpected adverse events to the study drug (ecopipam)[30]; the authors concluded that a run-in period would probably have prevented this.
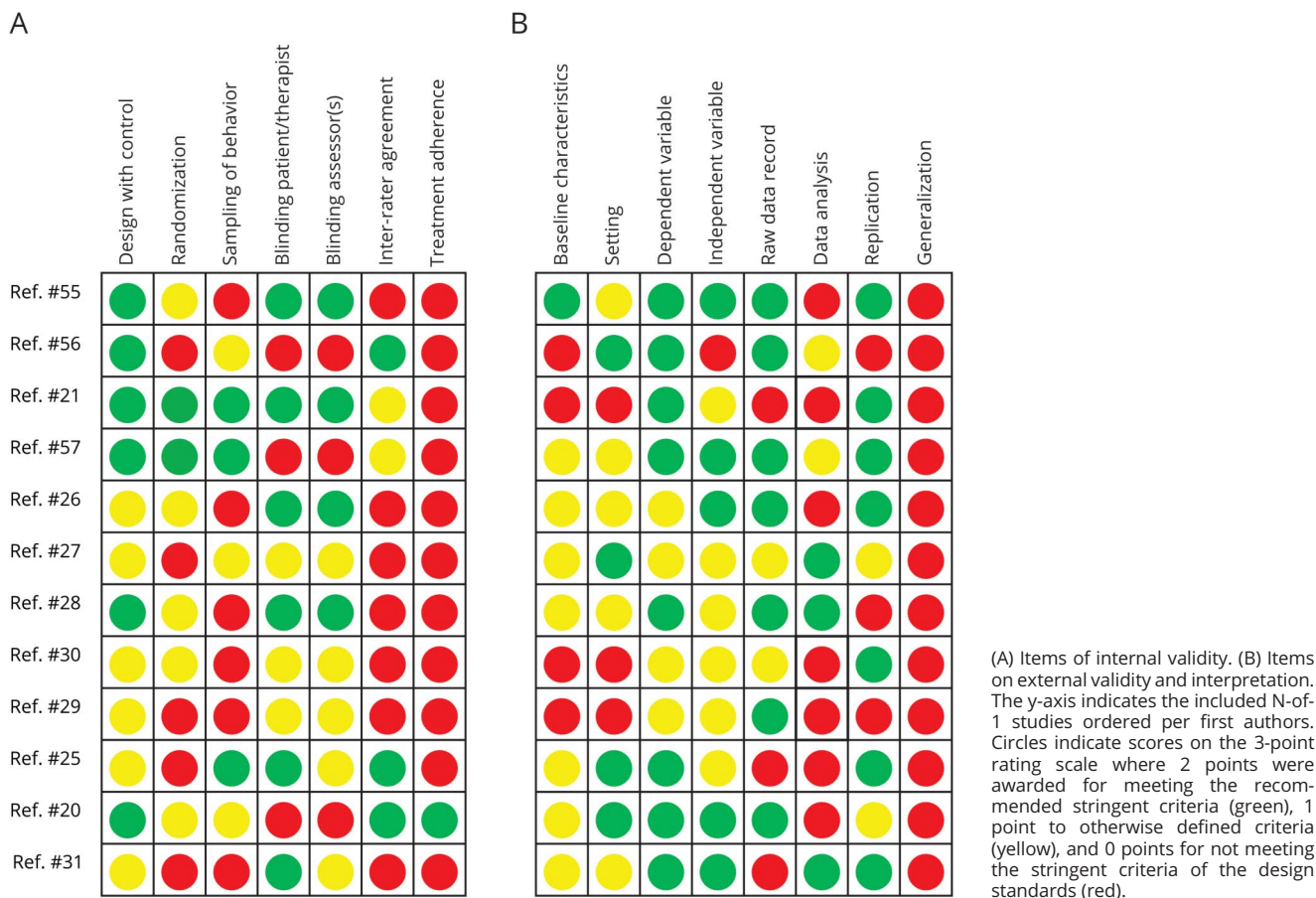
Suggested Inferences

In 9 studies, results were interpreted as generalizable to all patients with the same condition, whereas the authors of 2 studies considered the experiment as evidence for the individual participant only. One study did not report on inference.

Quality Assessment and Risk of Bias

Internal Validity

The median of the internal validity score of the included N-of-1 studies as assessed by the RoBiNT was 6.5 of 14 points

Figure 3 Schematic Representation of the Risk of Bias in N-of-1 Trials (RoBiNT) Scale

(A) Items of internal validity. (B) Items on external validity and interpretation. The y-axis indicates the included N-of-1 studies ordered per first authors. Circles indicate scores on the 3-point rating scale where 2 points were awarded for meeting the recommended stringent criteria (green), 1 point to otherwise defined criteria (yellow), and 0 points for not meeting the stringent criteria of the design standards (red).

(range 3–11; figure 3A). Treatment adherence was not assessed with the exception of 1 study that scored the maximum on treatment adherence by fulfilling the requirements of using a clear rating system, an independent assessor of the participant and sampling of more than 20% of the data, resulting in a minimum of 80% adherence.[20] The interrater agreement was adequately evaluated in 3 studies with separate reporting on the dependent variables for each condition.

External Validity and Interpretation
The median of the external validity score of the included N-of-1 studies was 9 of 16 points (range 4–11; figure 3B). The dependent variable (target behavior) was in 8 of 12 studies operationally defined with description of the measuring method. The other 4 studies did define the target behavior, but without clear and precise description of methods of measuring. Also, studies scored relatively high on describing practical matters including equipment, manuals, and procedural details. Although 1 study described the intervention in vague or general terms, 6 studies provided broad but not detailed descriptions of the content of the intervention or lacked one of the procedural's items including the number, duration, and frequency of periods for each participant. The other 5 studies provided a detailed description of the content

of the intervention, the procedure of delivery, and any equipment and manuals used. However, low scores were found on description of baseline characteristics (9/24 points), data analysis (8/24 points), and generalization (0/24 points), referring to the inclusion of generalization measures.

Reporting of the N-of-1 Trials Against CENT 2015 Criteria
None of the studies provided a registration number, name of trial registry, nor information about accessibility of the full trial protocol. Two studies identified the study as (a series of) N-of-1 trials in the title.[21,28] The rationale for using an N-of-1 approach was not clarified in any of the studies. Other omissions included the description and measurement properties including validity and reliability of outcome assessment tools, determination of sample size or requirement of the number of periods in a single N-of-1 study, and randomization and sequence allocation with a rationale or method. Carryover effects were not addressed, nor were period effects. As for the series, quantitative synthesis of individual data, including subgroup and sensitivity analyses, adjusted analyses, and analyses to determine heterogeneity between participants, were not reported. Moreover, (group) estimated effect sizes and its precision for each primary and secondary outcome were only reported in 2 studies.[27,31]

Figure 4 Challenges and Recommendations for Conducting and Reporting N-of-1 Studies in Rare Genetic Neurodevelopmental Disorders

	Challenges	Recommendations
Population	Heterogeneity	Clarify: - Disorder and diagnostic criteria - Baseline characteristics - Target symptom(s) - Eligibility - Concurrent therapies - Comorbid conditions Elaborate on setting and location
Intervention	Rationale	Distinguish disease-modifying and symptomatic intervention Take natural course into account Ensure optimal engagement with intervention and outcome measures
Design	Methodology	Substantiate duration of period based on pharmacokinetics and dynamics Appropriate number of crossover periods and valid off-periods Perform power analysis → Desired number of: - Participants - Periods - Frequency of measurements Appropriate randomization and sequences ensuring optimal blinding; deliberate on methods to use for randomization Consider adding: - Baseline period - Dose titration period - Run-in period (to ensure blinding, add to comparator period also) - Washout period considering biological and psychological effects - Follow-up measurement Acquire multiple data points per period
Outcome measures	Inference	Choose several levels of measurement, such as personalized, disorder-specific, biological, or neuropsychological Consider a measure for generalization
	Validity	Describe measurement properties, explicitly responsiveness to change, and applicability Elaborate on validity regarding the population (including disorder, intellectual disability, age)
	Compliance	Minimize burden for (proxy) raters Consider assessment by multiple raters in relevant settings
Analysis	Reliability	Perform: - (Non-)parametric analysis - Ancillary analyses - If possible, interim analyses to minimize burden to patient

Strengths of the N-of-1 Studies Identified

The main strengths reported by the studies' authors included individual-centered evidence-based interventions and the intent to measure personalized and clinically relevant outcomes. Other assets were independence of assessors, control for day-to-day variation in symptoms, and use of subjective as well as objective and biological measures of treatment. Reviewers identified additional strengths that were encountered in some but not all studies: proof of concept in relatively small studies, individual-centered, multiple assessors, inclusion of baseline conditions, (clinically) relevant outcome measures, inclusion of control participants to determine whether effect is specific to the genetic disorder, and the systematic approach.

Limitations of the N-of-1 Studies Identified

The authors of the conducted N-of-1 studies reported difficulty with identifying appropriate and validated outcome measures, especially for specific genetic heterogeneous conditions for which outcome measures were often subjective. Reviewers additionally identified unclear measurement properties as a limitation, involving reliability, validity, and responsiveness. Psychological interventions and outcome assessment were vulnerable to bias because of subjectivity, task engagement, and personal attention or interaction. In 1 study, indications for a strong negative caretaker bias of a seemingly already proven intervention based on anecdotal reports of efficacy and prejudices were reported to have

affected recruitment of participants, compliance, and, subsequently, outcome scores.[31] Also, a difference between ratings by caregivers and research personnel was perceived in some studies without assessing an interrater agreement. Finally, difficulty with statistical analysis was identified. As N-of-1 studies could have different purposes such as a proof of concept, providing an individual treatment decision, or estimating the treatment effect at a population level, the level of complexity and necessity of statistical analyses might be contingent on the reason for the study. Specifically, the degree of certainty desired was taken into consideration by the author(s) in 1 study where a visual analysis clearly showed that the active intervention was beneficial compared with placebo, but the statistical analyses did not reveal significant results in some cases.[28]

Discussion

N-of-1 studies have been recommended for evaluating the efficacy of interventions in rare disorders.[32,33] However, in this extensive review, only 12 studies complied with the fundamental N-of-1 criteria of a controlled multiple crossover trial, showing limited use and reporting of N-of-1 trials for rare genetic neurodevelopmental disorders. In addition to limitations in design and statistical analysis, generalizability and feasibility were particularly challenging. Below, limitations are discussed and recommendations are provided to implement and optimize future N-of-1 studies in this patient population (figure 4).

Although the genetic disorder and presence of ID were generally reported, diagnostic and eligibility criteria, comorbid conditions, and concurrent therapies were often unclear. Rare genetic neurodevelopmental disorders are often accompanied by various and often variable levels of ID and severe comorbidities. This intra- and interindividual heterogeneity can complicate generalization of findings to other patient populations. To optimize interpretation and generalizability, eligibility criteria and baseline characteristics pertaining to the study population as well as environment should be thoroughly described.

The rationale for the intervention was well described in the reported studies. Distinction between disorder-specific and disease-modifying drugs was not performed by the authors. Categorization was difficult for some included studies as interventions may be disorder specific and not directly change its natural course by for example not targeting the primary underlying pathophysiology as exemplified by the study to L-arginine supplementation in ornithine transcarbamylase deficiency.[28] Despite the fact that L-arginine supplementation does not target ornithine transcarbamylase itself but rather the consequences of the enzymatic deficiency, L-arginine ameliorates the overall function of the urea cycle by maintaining a normal rate of protein synthesis.[28]

To optimize impact of N-of-1 studies, it is important to specify whether a trial will focus on syndrome-specific or more common manifestations. Now that disease-modifying drugs are

becoming increasingly available,[18] consideration of disorder-specific effects is especially important with regard to generalizability to other patient populations. Also, disease-modifying drugs may have age- or comorbidity-dependent effects. For example, therapeutic effects of mammalian target of rapamycin inhibitors for tuberous sclerosis complex might differ over time, across patients, and across manifestations.[34,35] This emphasizes the need for detailed baseline characteristics.

The interventions of the included studies were mainly directed to neurobehavioral manifestations such as improving cognition, behavior, and quality of life, underlining the great burden of neuropsychiatric symptoms for patients as well as caregivers in patients with rare genetic neurodevelopmental disorders.[36] Considering the high burden of shared neuropsychiatric comorbidity, symptomatic interventions are of pivotal importance as their effect may be disorder transcending. Hence, especially symptomatic drug and nondrug trials should discuss generalizability of their intervention to other populations, taking disorder-specific effects and side effects into account. The critical need for well-controlled studies before interventions becomes established as standard of care was underscored by a negative caretaker bias encountered in 1 study.[31]

Only 2 studies were explicitly identified by the authors as an N-of-1 trial, underlining the need of a common terminology. The rationale for the N-of-1 design was generally not specified. Other limitations regarding the trial design were observed including unclear justification of trial and intervention duration, lack of run-in periods, carryover effects, randomization, and blinding.

It has been proposed that conditions should be stable over time to be eligible for conducting an N-of-1 study.[37] IEMs are however typically (neuro)degenerative disorders resulting in an unstable and often variable natural course across patients. As the natural history of other types of neurodevelopmental disorders unfolds, this variable course increasingly applies to many other genetic neurodevelopmental disorders.[38,39] However, even for unstable manifestations, effects may be observed by tracing the overall enduring effect on the personal course, including (multiple) baseline, placebo, and follow-up measurements. In this way, disease-modifying treatment options can be investigated, theoretically expecting a more enduring effect on the individual's natural course for disease-modifying drugs vs a temporary effect for symptomatic drug treatments.

To substantiate the duration of the interventional period, pharmacokinetics and dosage should be taken into full consideration. Dosage should be based on factors such as half-life time, age, weight, and daily timing. Both low dosages and high dosages without a run-in period can result in dropout and lack of efficacy.[21,30] Multiple dosages might be considered by implementing an ABC design or adjusting dosages after interim analyses.

To minimize carryover and side effects, addition of a run-in and/or washout period is preferred.[40,41] In addition to

biological carryover effects based on half-life time of drugs, psychological carryover effects for the patient as well as proxies should be considered, such as relief of parental stress after a period with an effective intervention. A baseline condition to observe natural behavior without any intervention and a follow-up will add internal validity and information about the effectiveness and tolerability of an intervention.

To gauge the robustness of methods chosen for randomization and sequence allocation, this should be thoroughly described, such as steps taken to conceal the sequence, information about who generated the sequence, who enrolled participants, and who assigned them to interventions. Various randomization and implementation methods may be appropriate depending on the condition and design.[40] Interpretation of observed effects becomes problematic with randomization when outcomes unexpectedly or progressively deteriorate or improve.[6] Counterbalancing can be used to systematically alternate the treatment order (such as ABBA instead of AABA or AABB) so that neither treatment suffers a worse fate than the other.[42]

In terms of personalized care, included studies were commendable by tailoring interventions to patient or caregiver needs, thus ensuring relevance and optimizing treatment adherence. Outcome measures included objective and biological outcomes, validated symptom checklists, neuropsychological assessments, or personalized outcomes. Preferably, all types are included to optimize pathophysiologic insights as well as relevance to the patient. Feasibility of N-of-1 studies in these vulnerable patients was questioned in 4 studies. As an N-of-1 study might be time and effort consuming for several stakeholders involved in the study because of frequent recording of data points enabling multiple measurements, and the number of periods and duration of the trial, increasing treatment adherence should be prioritized. To foster treatment and trial adherence, patient involvement on the intervention, design, and outcome measures appears to greatly contribute to the experienced relevancy and enthusiasm of participants.[43] However, this might strengthen potential placebo effects. As participants with ID can often not report on their clinical condition, this places a demand on parents and caregivers. Proxy-friendly assessment tools are required to ensure trial compliance.

Targeting behavioral outcomes in patients with rare disorders and varying levels of cognitive functioning is complex as appropriate outcome measures are limited and often lack validity.[44] Hence, interpretation of efficacy is hampered leading to disappointing results of disorder-specific interventional studies. This underlines the need for more sensitive and disorder-specific evaluation strategies, such as the phenylketonuria–quality of life (PKU-QOL) questionnaire.[45] For outcomes, the property responsiveness to change is essential in measuring the effectiveness of interventions but is often unknown. Of the included studies that used existing rating scales, responsiveness to change was discussed for Dementia Care Mapping, the Vineland Adaptive Behavior Scale, and the Behavior Problems Inventory.[46–48] Of interest to heterogeneous populations with

ID is the recently introduced NIH battery of neuro-psychological assessments, which is increasingly validated.[49]

As patients with rare genetic neurodevelopmental disorders comprise a vulnerable patient group often affected by severe comorbidity and complex environmental factors, there is a great need for personalized and disorder-specific outcome measures. This was also indicated by the frequent use of self-designed outcome measures in the included studies. Instruments such as patient-reported outcome measures,[44,50] Goal Attainment Scaling,[51] or experience-sampling methods[52] may be considered, enabling quantitative expression of meaningful subjective patient experiences while translating these into evidence.[43] As personalized outcome measures may compromise generalizability, inclusion of generalization measures can provide information on transfer effects of the intervention to other behaviors, settings, or disorders that may be either closely or distally related to the target behavior.[15]

One main shared shortcoming was the lack of statistical analyses. None of the 12 studies included a justification for the sample size. Sample size calculations are important to ensure that clinically relevant effects can be detected while not including, and hence burdening, too many patients. In N-of-1 trials, a power analysis can help decide on the number of periods required to detect a clinically relevant treatment effect within a patient and, in case of a series of N-of-1 trials, for the number of participants required to determine an average treatment effect in the study sample. Formulas and methods for calculation of the required sample size for these different objectives are available for N-of-1 studies.[53]

The majority of the studies only described results using graphical or tabular methods, whereas (non)parametric statistical analyses are now considered the standard for testing for an intervention effect in N-of-1 studies.[54] (Non)parametric and ancillary analyses should be performed to evaluate period effects, intrasubject correlations, and subgroup and adjusted effects. Rather than attempting to adjust for carryover effects, it is preferred to choose the (washout) periods long enough for carryover not to occur.

Both mixed-effects models and Bayesian models can properly address the inter- and intrapatient variability in series of N-of-1 trials.[41] A clear overview is given of the various frequentist analyses proposed for N-of-1 trials that may serve different purposes.[40] Most importantly, the statistical methods should properly account for the method of randomization used. Simple analyses such as a paired t test and a summary measure approach can be acceptable for testing the hypothesis of a difference between treatments. For assessing heterogeneity of the treatment effects between individuals, a mixed model approach is required[40] with an ANOVA type test for hypothesis testing. The latter can also be done in a Bayesian framework using hierarchical modeling. In a Bayesian framework, it is quite natural to update an estimation when data from new N-of-1 trials become available. If one wishes to

produce shrunken estimates or predict the effects for future patients, a hierarchical Bayesian model or linear mixed model with random treatment-by-patient interaction is required.

Reporting an N-of-1 trial should satisfy particular N-of-1 items according to CENT 2015 and RoBiNT[15,23] (figure 4). Because of the differences in N-of-1 terminology that still exist, studies should identify the trial as an N-of-1 in both the title and the abstract. In addition to the items discussed above and in Figure 4, a rationale for using the N-of-1 design should be provided because N-of-1 trials may serve a number of different purposes[53] and several single-case experimental designs could be considered.[32] More specifically, we especially recommend an N-of-1 study in rare genetic disorders when the intervention has a predictable duration of effect for which a valid off-period is possible and low recruitment rates are expected. Finally, trial registration and an accessible full trial protocol including specific methodological choices might be of pivotal importance for future N-of-1 studies. In line with recent guidelines for N-of-1 trial protocols and reporting,[6,8] we recommend facilitation of entry of N-of-1 studies into primary registries within the World Health Organization's International Trials Registry Network and clinicaltrials.gov.

A strength of this study is the comprehensive search strategy necessitated by the historical lack of uniformity in N-of-1 terminology. The large amount of records identified through this search inadvertently may have resulted in inappropriate exclusions. N-of-1 studies that were directed toward symptoms solely without mentioning underlying disorders might also have been missed as our search was developed with a gene first approach. Of note, the recommendations reflect the authors' opinions rather than a systematically derived consensus.

N-of-1 studies have great potential to provide evidence of effectiveness for individuals as well as groups of patients. The findings of this review show only limited use of N-of-1 studies in rare genetic neurodevelopmental disorders and that improvement of methodology is essential to provide a suitable alternative for RCTs. We provide recommendations to enhance methodological and statistical quality as well as generalizability, feasibility, and personalization. Future use of this N-of-1 framework will assist in realizing the sorely needed evidence-based interventions for these vulnerable patients.

Study Funding

Disclosure

Publication History

Appendix Authors

Name	Location	Contribution
Annelieke R Müller, MSc	's Heeren Loo, Amersfoort, the Netherlands; Amsterdam UMC, Pediatric Metabolic Diseases, Emma Children's Hospital, University of Amsterdam, Amsterdam, the Netherlands	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design; and analysis or interpretation of data
Marion MMG Brands, MD, PhD	Amsterdam UMC, Pediatric Metabolic Diseases, Emma Children's Hospital, University of Amsterdam, Amsterdam, the Netherlands	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; and study concept or design
Peter M van de Ven, PhD	Amsterdam UMC, Department of Epidemiology and Biostatistics, Amsterdam, the Netherlands	Major role in the acquisition of data and study concept or design
Kit CB Roes, Prof, PhD	Radboud University Medical Center, Department of Health Evidence, Biostatistics, Radboud University, Nijmegen, the Netherlands	Study concept or design and analysis or interpretation of data
Martina C Cornel, Prof, MD, PhD	Amsterdam UMC, Vrije Universiteit Amsterdam, Department of Clinical Genetics, Amsterdam Public Health research institute, Amsterdam, the Netherlands	Major role in the acquisition of data; study concept or design; and analysis or interpretation of data
Clara DM van Karnebeek, Prof, MD, PhD	Amsterdam UMC, Pediatric Metabolic Diseases, Emma Children's Hospital, University of Amsterdam, Amsterdam, the Netherlands; Radboud University Medical Center, Department of Pediatrics, Radboud Centre for Mitochondrial Medicine, Nijmegen, the Netherlands	Study concept or design
Frits A Wijburg, Prof, MD, PhD	Amsterdam UMC, Pediatric Metabolic Diseases, Emma Children's Hospital, University of Amsterdam, Amsterdam, the Netherlands	Drafting/revision of the manuscript for content, including medical writing for content, and study concept or design
Joost G Daams, MA	Amsterdam UMC, Medical Library, University of Amsterdam, Amsterdam, the Netherlands	Analysis or interpretation of data; additional contributions: search strategy
Erik Boot, MD, PhD	's Heeren Loo, Amersfoort, the Netherlands; Maastricht University, Department of Psychiatry & Neuropsychology, Maastricht, the Netherlands, University Health Network, The Dalglish Family 22q Clinic, Toronto, Ontario, Canada	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design; and analysis or interpretation of data

Continued

Appendix (continued)

Name	Location	Contribution
Agnies M van Eeghen, MD, PhD	's Heeren Loo, Amersfoort, the Netherlands; Amsterdam UMC, Emma Children's Hospital, University of Amsterdam, Amsterdam, the Netherlands; Erasmus Medical Center, ENCORE, Rotterdam, the Netherlands	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design; and analysis or interpretation of data

References

1. European Parliament C of the EU. Regulation (EC) N°141/2000 of the European Parliament and of the Council of 16 December 1999 on Orphan Medicinal Products 2000. Official Journal of the European Communities. Available at: ec.europa.eu/health/sites/health/files/files/eudralex/vol-1/reg_2000_141_cons-2009-07/reg_2000_141_cons-2009-07_en.pdf. Accessed February 20, 2020.
2. European Organisation for Rare Diseases (EURORDIS): Rare Diseases: Understanding This Public Health Priority. 2005. Available at: eurordis.org/IMG/pdf/princeps_document-EN.pdf. Accessed February 20, 2020.
3. Morant AV, Jagalski V, Vestergaard HT. Labeling of disease-modifying therapies for neurodegenerative disorders. Front Med 2019;6:223.
4. Vernon HJ. Inborn errors of metabolism: advances in diagnosis and therapy. JAMA Pediatr 2015;169:778–782.
5. Guyatt G, Sackett D, Adachi J, et al. A clinician's guide for conducting randomized trials in individual patients. Can Med Assoc J 1988;139:497.
6. Shamseer L, Sampson M, Bukutu C, et al. CONSORT extension for reporting N-of-1 trials (CENT) 2015: explanation and elaboration. BMJ 2015;350:1793.
7. OCEBM Levels of Evidence Working Group. The Oxford Levels of Evidence 2. Oxford Centre for Evidence-Based Medicine; 2011. Available at: cebm.net/index.aspx?o=5653. Accessed February 20, 2020.
8. Porcino AJ, Shamseer L, Chan AW, et al. SPIRIT extension and elaboration for n-of-1 trials: SPENT 2019 checklist. BMJ 2020:368.
9. Lamont A, Lyons MD, Jaki T, et al. Identification of predicted individual treatment effects in randomized clinical trials. Stat Methods Med Res 2018;27:142–157.
10. Mitchell GK, Hardy JR, Nikles CJ, et al. The effect of methylphenidate on fatigue in advanced cancer: an aggregated N-of-1 trial. J Pain Symptom Manag 2015;50:289–296.
11. Blackston J, Chapple A, McGree J, McDonald S, Nikles J. Comparison of aggregated N-of-1 trials with parallel and crossover randomized controlled trials using simulation studies. Healthcare 2019;7:137.
12. Bradbury J, Avila C, Grace S. Practice-based research in complementary medicine: could N-of-1 trials become the new gold standard? Healthcare 2020;8:15.
13. Schork NJ. Personalized medicine: time for one-person trials. Nature 2015;520:609–611.
14. Goetz LH, Schork NJ. Personalized medicine: motivation, challenges, and progress. Fertil Sterility 2018;109:952–963.
15. Tate RL, Perdices M, Rosenkoetter U, et al. Revision of a method quality rating scale for single-case experimental designs and n-of-1 trials: the 15-item Risk of Bias in N-of-1 Trials (RoBiNT) Scale. Neuropsychol Rehabil 2013;23:619–638.
16. Porcino AJ, Punja S, Chan AW, et al. Protocol for a systematic review of N-of-1 trial protocol guidelines and protocol reporting guidelines. Syst Rev 2017;6:1–5.
17. Moher D, Shamseer L, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2016;4:1.
18. Van Karnebeek CDM, Stockler S. Treatable inborn errors of metabolism causing intellectual disability: a systematic literature review. Mol Genet Metab 2012;105:368–381.
19. Van Karnebeek CDM, Shevell M, Zschocke J, Moeschler JB, 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.
20. Simacek J, Dimian AF, McComas JJ. Communication intervention for young children with severe neurodevelopmental disabilities via telehealth. J Autism Dev Disord 2017;47:744–767.
21. Camfield P, Gordon K, Dooley J, Camfield C. Melatonin appears ineffective in children with intellectual deficits and fragmented sleep: six "N of 1" trials. J Child Neurol 1996;11:341–343.
22. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016;5:210.
23. Vohra S, Shamseer L, Sampson M, et al. CONSORT extension for reporting N-of-1 trials (CENT) 2015 Statement. BMJ 2016;350:h1738.
24. Cummings JL. Defining and labeling disease-modifying treatments for Alzheimer's disease. Alzheimer's Demen 2009;5:406–418.
25. Marholin D, Pohl RE, Stewart RM, Touchette PE, Townsend NM, Kolodny EH. Effects of diet and behavior therapy on social and motor behavior of retarded phenylketonuric adults: an experimental analysis. Pediatr Res 1978;12:179–187.
26. Fisch GS, Cohen IL, Gross AC, Jenkins V, Jenkins EC, Brown WT. Folic acid treatment of fragile X males: a further study. Am J Med Genet 1988;30:393–399.
27. Giffin FD, Clarke JTR, D'entremont DM. Effect of dietary phenylalanine restriction on visual attention span in mentally retarded subjects with phenylketonuria. Can J Neurol Sci 1980;7:127–131.
28. Hackett A, Gillard J, Wilcken B. N of 1 trial for an ornithine transcarbamylase deficiency carrier. Mol Genet Metab 2008;94:157–161.
29. Luciano AY, Jinnah HA, Pfeiffer RF, Truong DD, Nance MA, LeDoux MS. Treatment of myoclonus-dystonia syndrome with tetrabenazine. Parkinsonism Relat Disord 2014;20:1423–1426.
30. Khasnavis T, Torres RJ, Sommerfeld B, Puig JG, Chipkin R, Jinnah HA. A double-blind, placebo-controlled, crossover trial of the selective dopamine D1 receptor antagonist ecopipam in patients with Lesch-Nyhan disease. Mol Genet Metab 2016;118:160–166.
31. Tierney E, Conley SK, Goodwin H, Porter FD. Analysis of short-term behavioral effects of dietary cholesterol supplementation in Smith-Lemli-Opitz syndrome. Am J Med Genet A 2010;152:91–95.
32. Gupta S, Faughnan ME, Tomlinson GA, Bayoumi AM. A framework for applying unfamiliar trial designs in studies of rare diseases. J Clin Epidemiol 2011;64:1085–1094.
33. Guyatt GH, Haynes RB, Jaeschke RZ, et al. Users' guides to the medical literature: XXV. Evidence-based medicine: principles for applying the users' guides to patient care. J Am Med Assoc 2000;284:1290–1295.
34. Curatolo P, Moavero R. mTOR inhibitors in tuberous sclerosis complex. Curr Neuropharmacol 2012;10:404–415.
35. Shapiro EG, Escolar ML, Delaney KA, Mitchell JJ. Assessments of neurocognitive and behavioral function in the mucopolysaccharidoses. Mol Genet Metab 2017;122:8–16.
36. Jansen DEMC, Krol B, Groothoff JW, Post D. People with intellectual disability and their health problems: a review of comparative studies. J Intellect Disabil Res 2004;48:93–102.
37. Nikles J, Mitchell GK, Schluter P, et al. Aggregating single patient (n-of-1) trials in populations where recruitment and retention was difficult: the case of palliative care. J Clin Epidemiol 2011;64:471–480.
38. Ballard C, Mobley W, Hardy J, Williams G, Corbett A. Dementia in Down's syndrome. Lancet Neurol 2016;15:622–636.
39. Van Eeghen AM, Black ME, Pulsifer MB, Kwiatkowski DJ, Thiele EA. Genotype and cognitive phenotype of patients with tuberous sclerosis complex. Eur J Hum Genet 2012;20:510–515.
40. Araujo A, Julious S, Senn S. Understanding variation in sets of N-of-1 trials. PLoS One 2016;11:e0167167.
41. Senn S. Mastering variation: variance components and personalised medicine. Stat Med 2016;35:966–977.
42. Schmid CH, Duan N. Statistical design and analytic considerations for N-of-1 trials. In: Design and Implementation of N-Of-1 Trials: A User's Guide. Rockville: Agency for Healthcare Research and Quality (AHRQ); 2014:33–53.
43. 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 2018;122:1287–1294.
44. Slade A, Isa F, Kyte D, et al. Patient reported outcome measures in rare diseases: a narrative review. Orphanet J rare Dis 2018;13:1–9.
45. Regnault A, Burlina A, Cunningham A, et al. Development and psychometric validation of measures to assess the impact of phenylketonuria and its dietary treatment on patients' and parents' quality of life: the phenylketonuria–quality of life (PKU-QOL) questionnaires. Orphanet J rare Dis 2015;10:59.
46. Ettema TP, Dröes RM, Lange JDe, Mellenbergh GJ, Ribbe MW. A review of quality of life instruments used in dementia. Qual Life Res 2005;14:675–686.
47. Harris SL, Handleman JS, Belchic J, Glasberg B. The vineland adaptive behavior scales for young children with autism. Spec Serv Schools 1995;10:45–54.
48. Rojahn J, Schroeder SR, Mayo-Ortega L, et al. Validity and reliability of the Behavior Problems Inventory, the Aberrant Behavior Checklist, and the Repetitive Behavior Scale - revised among infants and toddlers at risk for intellectual or developmental disabilities: a multi-method assessment approach. Res Dev disabilities 2013;34:1804–1814.
49. Shields RH, Kaat AJ, McKenzie FJ, et al. Validation of the NIH toolbox cognitive battery in intellectual disability. Neurology 2020;94:1229–1240.
50. O'Mahony B, Skinner MW, Noone D, Page D, O'Hara J. Assessments of outcome in haemophilia—a patient perspective. Haemophilia 2016;22:208–209.
51. Gaasterland CMW, Van Der Weide MCJ, Roes KCB, Van Der Lee JH. Goal attainment scaling as an outcome measure in rare disease trials: a conceptual proposal for validation. BMC Med Res Methodol 2019;19:227.
52. Csikszentmihalyi M, Larson R. Validity and reliability of the experience-sampling method. In: Flow and the Foundations of Positive Psychology. Dordrecht: Springer; 2014:35–54.
53. Senn S. Sample size considerations for n-of-1 trials. Stat Methods Med Res 2019;28:372–383.
54. Craig P, Dieppe P, Macintyre S, Mitchie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008;337:a1655.
55. Bawden HN, MacDonald GW, Shea S. Treatment of children with Williams syndrome with methylphenidate. J Child Neurol 1997;12:248–252.
56. Byiers BJ, Dimian A, Symons FJ. Functional communication training in Rett syndrome: a preliminary study. American J Intellect Dev Disabilities 2014;119:340–350.
57. Crook N, Adams M, Shorten N, Langdon PE. Does the well-being of individuals with down syndrome and dementia improve when using life story books and rummage boxes? A randomized single case series experiment. J Appl Res Intellect Disabilities 2016;29:1–10.