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

Neurology® 2021;96:529-540. doi:10.1212/WNL.0000000000011597

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.

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

RELATED ARTICLE

Editorial
N-of-1 Trials in Rare Genetic Neurodevelopmental Disorders: Opportunities for Improvement

Page 513
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 searching (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
reporting N-of-1 Trials (CENT) 2015 and the Risk of Bias in N-of-1 Trials (RoBiNT) Scale15 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

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adherence</td>
<td>The extent to which a participant’s behavior matches agreed recommendations from a health care provider taking into account the patient’s perspectives.</td>
</tr>
<tr>
<td>Block</td>
<td>A repeated unit of a set number of periods.</td>
</tr>
<tr>
<td>Compliance</td>
<td>The extent to which the patient’s behavior corresponds with the prescriber’s recommendations.</td>
</tr>
<tr>
<td>Cycle</td>
<td>Each repeated unit of a set number of periods within a sequence (e.g., ABA).</td>
</tr>
<tr>
<td>Generalization</td>
<td>The degree to which results observed in a study may extend to other patients or settings, providing an indication of the external validity.</td>
</tr>
<tr>
<td>Generalization measure</td>
<td>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.</td>
</tr>
<tr>
<td>Internal validity</td>
<td>The degree to which the study’s outcomes could be attributed to the intervention being responsible for change in the dependent variable.</td>
</tr>
<tr>
<td>N-of-1 study</td>
<td>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.</td>
</tr>
<tr>
<td>Pair</td>
<td>A repeated unit containing only 2 periods.</td>
</tr>
<tr>
<td>Period</td>
<td>The duration of an intervention, comparator, washout, or run-in.</td>
</tr>
<tr>
<td>Suggested inference</td>
<td>Interpretation of the extent of generalization of the study’s outcomes to either the individual participants or patients in general with that specific disorder.</td>
</tr>
<tr>
<td>Responsiveness to change</td>
<td>The ability of an instrument to detect change over time in a construct being measured.</td>
</tr>
<tr>
<td>Run-in</td>
<td>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.</td>
</tr>
<tr>
<td>Washout</td>
<td>Time without an intervention following a treatment period to ensure that effects of treatment have disappeared.</td>
</tr>
</tbody>
</table>
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.¹⁸ 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.
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 l-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

### 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.
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. Once, a quality of life assessment was used. 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); 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
(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. 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. 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.
**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.\textsuperscript{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.\textsuperscript{28}

\section*{Discussion}

N-of-1 studies have been recommended for evaluating the efficacy of interventions in rare disorders.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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,\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{37}

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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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 pathophysiological 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 relevance 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 neuropsychological 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 methods52 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 distantly 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 required40 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 RoBiNT15,23 (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 purposes53 and several single-case experimental designs could be considered.52 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**
No targeted funding reported.

**Disclosure**
The authors report no disclosures relevant to the manuscript. Go to Neurology.org/N for full disclosures.

**Publication History**
Received by *Neurology* August 11, 2020. Accepted in final form December 18, 2020.

---

### Appendix Authors

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


41. Wolfr C, Jo, Heeren Loo, Maersk, T., Vogel, K., Gordon K, Dooley J, Cam...
Systematic Review of N-of-1 Studies in Rare Genetic Neurodevelopmental Disorders: The Power of 1
Annelieke R. Müller, Marion M.M.G. Brands, Peter M. van de Ven, et al.
Neurology 2021;96:529-540 Published Online before print January 27, 2021
DOI 10.1212/WNL.0000000000011597

This information is current as of January 27, 2021

Updated Information & Services
including high resolution figures, can be found at:
http://n.neurology.org/content/96/11/529.full

References
This article cites 51 articles, 2 of which you can access for free at:
http://n.neurology.org/content/96/11/529.full#ref-list-1

Citations
This article has been cited by 3 HighWire-hosted articles:
http://n.neurology.org/content/96/11/529.full##otherarticles

Subspecialty Collections
This article, along with others on similar topics, appears in the following collection(s):
All Genetics
http://n.neurology.org/cgi/collection/all_genetics

Permissions & Licensing
Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at:
http://www.neurology.org/about/about_the_journal#permissions

Reprints
Information about ordering reprints can be found online:
http://n.neurology.org/subscribers/advertise